A noticing-promotion approach and L2 development:

a study of English interrogative acquisition in the classroom

Atsuko Moriyama

Ph.D.

University of Edinburgh

2006

ABSTRACT

Instruction which draws learner attention to form in meaning-centred activities is often referred to as *focus on form* (FonF) (Long, 1991). It is considered a potentially useful means of helping learners to acquire L2 forms, since it can promote the *noticing* of linguistic gaps between their output and the target language, which is regarded as a necessary condition for L2 development (Schmidt, 1994a). Among the most extensively studied L2 features are English interrogatives (Mackey, 1999; Mackey & Oliver, 2002; Mackey & Philp, 1998; McDonough, 2005; Silver, 1999, 2000; Spada & Lightbown, 1993). These studies have shown the benefit of FonF instruction in learners' advancement in developmental stage, but most research to date has been conducted in experimental conditions and has focused mainly on the effect of intensive corrective feedback on learners' nontarget-like L2 use by a native speaker in one-to-one conversations; few attempts have been made to examine effective ways of implementing FonF instruction in typically large EFL classes, where such feedback is difficult to provide.

This study aims to fill this gap and explores the potential of teachers' creating opportunities for *noticing* the gap through dialogical interaction between learners. To this end, a pedagogical option, a *noticing-promotion approach*, was employed, intended to encourage learners to take an active role both in *noticing* gaps and in assisting their partner by scaffolding replies to notice and self-correct mistakes. The instruction consisted of four sessions, including explicit teaching of one or two pre-selected structures and noticing tasks (dictogloss and information-gap) with a modelling video performed by a learner dyad; this approach, therefore, represents a compromise between FonF and traditional presentation-practice-production instruction, with a strong leaning towards the latter, though not as far as FonF*S* as defined by Long (1991). A pretest and two posttests design was used to measure the sustained effect over 6-7 weeks. Oral performance data collected in two communicative tasks from 48 Japanese students was used to examine individuals' change in developmental stage in relation to their readiness to learn the target rules; the results were compared with those from a comparison group (N=12). In addition, the study explored through questionnaires learners' perceptions of their gains and of the usefulness of this type of instruction.

Results indicate that this *noticing-promotion approach* was effective in helping learners who were ready to learn the target rule advance in developmental stage and to fill gaps in their L2 development. Moreover, learners' self-reports showed that the instruction was beneficial, not only in motivating many ready students to carry on using the target rules beyond the treatment sessions, but also in helping them become better 'noticers' and users of the target rules both inside and outside the classroom.

DECLARATION

In accordance with Regulation 3.8.7 of the programme of postgraduate study, I declare that this thesis has been composed entirely by myself. This work it contains is my own and has not been submitted for any other degree or professional qualification.

Atsuko Moriyama

Edinburgh,

December 2006

Miyoko Moriyama
in memoriam

ACKNOWLEDGEMENTS

In the course of writing this thesis, numerous people have supported me, directly or indirectly. Without their help, I would not have been able to complete it.

First of all, I would like to express my heartfelt gratitude to my three supervisors: Tony Lynch, my principal supervisor, and Joan Maclean and Cathy Benson, my second supervisors. Tony has been supervising me since I was an MSc student and his critical and rapid feedback always supported and encouraged me. His wit and patience always made our meetings a pleasure, which kept me on the path to PhD completion. Joan provided me with useful comments on my teaching and testing materials. Cathy took over from Joan upon her retirement and Cathy's support and encouragement has been a source of energy in the long journey, especially when things were most difficult when I twice needed to suspend my study.

My sincere thanks also go to Professor Toyoko Shimizu, who offered me an opportunity to carry out my main study with her students at Chiba University. Without her support and encouragement, this thesis would never have been completed. I would also thank my colleagues at Kawatetsu Nursing College who offered unstinting moral support, particularly Kayoko Matsubara and Chigusa Akiyama.

I am deeply indebted to four organisations for their financial support during my PhD study. Firstly, my sincere thanks go to Universities UK (The Overseas Research Students Awards Scheme) and The University of Edinburgh Faculty of Arts (Faculty Group Research Scholarship), who jointly supported my PhD study. Thanks also go to the University of Edinburgh Development Trust, who supported me with the Small Project Grant Scheme. In addition, I am grateful to Rotary International, especially the Chiba-Minato Rotary Club in Japan and the Corstorphine Rotary Club in Edinburgh, who supported my MSc study, which was a stepping stone to my PhD study. I owe particular thanks to Mike Galloway, of the latter club, who was my advisor, and to his wife Heather, who have extended support to me ever since.

I would like to thank many friends both in Japan and in Edinburgh who supported my work. Kei Harada, Mitsuyo Takeuchi, Akiko Majima, and Makiko Saito helped me to administer three tests in the main study. Without their help, I would have been unable to collect such a large amount of data. Mikiko Kurokawa and Keiko Kageyama helped me to prepare two versions of the modelling video for my teaching. The recording extended till late at night several times, but they never showed signs of fatigue and shared their interest with me. I am most grateful to Aileen Little, who patiently proof-read my drafts, and to Kelly Peake and Melissa Macphail, who proof-read the course materials and an early version of my drafts. In addition, my sincere thanks go to Miyoko Kobayashi for her constant encouragement and useful comments by sharing her experiences as a former PhD student in Britain.

I would also like to express my special thanks to the students who participated in the data collection. It was through observing their learning process – struggle in learning the target structures, the glitter in their eyes when something clicked in their mind, and effort to assist each other in pair work – that I have learnt a great deal. At the same time, our shared experiences as learners of English kept motivating me

to complete this thesis.

My sincere thanks also go to Dr. Patsy M. Lightbown and Dr. Jennifer Philp for their prompt email replies to my questions about their coding question examples and for valuable comments.

My heartfelt thanks also go to José for his interest in my work and unfailing encouragement and support. The time we spent together in our kitchen during our spare moments and its results were nutritious to both my soul and body.

Lastly, I am most grateful to my mother, to whose memory this thesis is dedicated, and to my father, with whom she now rests, for their love, which kept nurturing me all these years.

CONTENTS

CHAPTER 1

CHAPTER 2

CHAPTER 3

CHAPTER 8

Learners' perceptions of the usefulness of the treatment 238

CHAPTER 9

APPENDICES

TABLES AND FIGURES

TABLES

FIGURES

ABBREVIATIONS

Affirmative-Aux/Do-2nd	A stage 5 question form. In *wh/Q*-word questions, an affirmative form of an auxiliary verb is placed in second position to a *wh/Q*-word before a subject followed by a main verb.
ALT	Assistant Language Teachers
Cancel-inversion	A stage 6 question form. In indirect questions, there is no inversion of the subject and the auxiliary/verb in the subordinate clause.
CG	The Comparison Group.
CLT	Communicative Language Teaching.
CR	Consciousness-raising.
EFL	English as a foreign language.
ESL	English as a second language.
FonF	Focus on form.
FonF*S*	Focus on form*S*.
Fronting	A stage 3 question form. In *yes/no* questions, the auxiliary 'do', or other lexical elements such as a *wh-/Q*-word or an auxiliary verb other than 'do', is placed in sentence-initial position in direct questions.
JET programme	Japan Exchange and Teaching programme.
IL	Interlanguage.
L1	A learner's first/mother language.
L2	A foreign or second language.
LREs	Language-related episodes.
MEXT	Ministry of Education, Culture, Sports, Science and Technology in Japan.
Negative-Aux/Do-2nd	A stage 5 question form. In *wh*-questions, a negated form of an auxiliary verb is placed in second position to a *wh/Q*-word before a subject followed by a main verb.
NNS	Non-native speakers.
NPAH	The Noun Phrase Accessibility Hierarchy.
NS	Native speakers.
P-P-P	Presentation-practice-production.
Q-tag	A stage 6 question form. In tag questions the following requirements are realised: (a) an affirmative form of an auxiliary verb should be present in the question tag if a negated form of an auxiliary verb is used in the main clause, and vice versa; (b) there should be person agreement between the subject of the sentence and a pronoun in the question tag; and (c) there should be number agreement between an auxiliary/main verb of the sentence and an auxiliary verb in the question tag.
Q-form	Question form
SLA	Second language acquisition
Stage 6-Ready	The students who had reached stage 5 in developmental stage in English interrogative acquisition on the pretest in this study.
Stage 6-Unready	The students who had not reached stage 5 in developmental stage in English interrogative acquisition on the pretest in this study.
TBI	Task-based instruction.
TG	Treatment Group.
TL	Target language.
TOEFL	Test of English as a Foreign Language.
Wh-inversion	A stage 4 question form. *Wh*-questions in which the copula is placed before the subject.
Y/N-Negative	A stage 6 question form. In *yes/no*-questions, a negated form of auxiliary verb is placed in the initial-position of the sentence before the subject.
Y/N-inversion	A stage 4 question form. *Yes/no*-questions in which an affirmative form of auxiliary verb other than 'do' is placed in sentence-initial position before the subject.
ZISA project	Zweitspracherwerb Italienischer und Spanischer Arbeiter project.
ZPD	Zone of proximal development.

CHAPTER 1
Introduction

In this study I attempt to address the needs of learners who are obliged to study English in large classes, where immediate teacher feedback on individuals' mistakes and errors[1] is difficult to provide. I also attempt to examine the effects of instruction, which draws learners' attention to *form* in communicatively oriented classrooms, on their advancement in developmental stage in the acquisition of interrogatives. Within these larger goals, I also explore two aspects of L2 learning:

(a) the potential and limitations of teachers' creating opportunities for *noticing the gap* through dialogical interaction between learners and

(b) learners' perceptions of their gains and usefulness of this type of instruction, which encourage them to help each other through scaffolding replies.

In Section 1.1 I will briefly introduce the background to my study, the setting, and the motivations for carrying it out. Firstly, in order to contextualise this study, I will describe the nature and problems of English language teaching in formal education in Japan, where I carried out my study. I will then discuss various constraints on implementing a communicative methodology in the classroom in Japan. Then I will discuss a need for classroom-based studies which take into account the constraints and the requirements of Japanese students and teachers in adopting a communicative methodology in the classroom. In Section 1.2, I will outline the structure of the remaining chapters of my thesis.

1.1 Contextualization of the present study

1.1.1 The Japanese context

In Japan almost all students begin to learn English as a subject on entering secondary school (age 12)[2]. Almost all English courses there are taught by Japanese teachers, although team-teaching, with an assistant language teacher (ALT) who is a native English speaker, has also been practised in some

English classes. ALTs are hired by the Ministry of Education, Culture, Sports, Science and Technology (MEXT), through the JET (Japan Exchange and Teaching) programme, which aims to promote communication in English for purposes of international understanding. However, ALTs are not allowed to teach on their own because they very rarely possess a certificate for teaching in Japan and many have no experience in English teaching.

Since the introduction of ALTs, opportunities to use English in the classroom have increased slightly, but in everyday life in Japan there is still little opportunity to use English for communication either in school or out of it, as is often the case in the EFL context, and explicit teaching of the elements of the language system has been widely practised by teachers. However, demand for attainment of adequate communicative competence in English is high, since it is considered to be 'a key to boosting Japan's diminishing competitiveness in the international market, and one of the important educational demands is that citizens possess adequate "linguistic (i.e. FL and primarily English) abilities, power to convey their own thoughts to others in a multinational environment, understand different cultures, and argue convincingly" (Nakaura, 2002:15)' (Taguchi, 2002:6). In order to meet this demand, the National Curricula, commonly called "Course of Study", for foreign language courses for lower and upper secondary school were revised by MEXT[3] in 1988 and the "communicative" Course of Study was introduced. This new curriculum features explicitly communicative goals with a balanced focus on all four skills, and introduced new Oral Communication courses into the upper secondary curriculum. Since then, MEXT has made further revisions, but the basic goal set in the 1988 version remains unchanged. For example, in the current Course of Study for lower and upper secondary school, the overall objectives of foreign language courses are given as follows:

The Course of Study for Lower Secondary School (implemented in 2002)

> To develop students' basic practical communication abilities such as listening and speaking, deepening the understanding of language and culture, and fostering a positive attitude toward communication through foreign languages. (MEXT, 2006. URL: www.mext.go.jp/english/shotou/030301.htm)

The Course of Study for Upper Secondary School (implemented in 2003)

> To develop students' practical communication abilities such as understanding information and the speaker's or writer's intentions, and expressing their own ideas, deepening the understanding of language and culture, and fostering a positive attitude toward communication through foreign

languages. (MEXT, 2006. URL: www.mext.go.jp/english/shotou/030301.htm)

The revision of the Course of Study was influenced by theories of applied linguistics developed in Europe and the United States of America and, on the introduction of the "communicative" curriculum, grammar courses were discarded. Since then, the integration of grammar teaching with a communication-oriented instruction has been a primary concern among L2 teachers. Teachers have been urged to change their practice from a traditional to a communication-centred approach, but they are concerned about both the effectiveness of a communicative methodology in helping learners acquire L2 structures and the outcomes of interaction between learners with limited L2 proficiency. I share these concerns from my five years' experience as a non-native teacher of English at three different types of institution in Japan and this was the starting point of my study.

It has been 18 years since the introduction of the "communicative" curriculum, but it is often pointed out there is a wide gap between what MEXT aims to accomplish and classroom practice (McConell, 2000; Ryan & Makorova, 2004; Smith & Imura, 2004; Taguchi, 2002; Taguchi, 2002; Wada, 1989). For example, one serious concern is the separation of grammar teaching and communication. The failure of Japanese learners to achieve a high level of communicative competence even after the six years study at secondary schools has been ascribed to a number of reasons. The most frequently suggested cause is MEXT's failure to take account of various constraints to implementing the imported methodology in the local context (McConell, 2000; Ryan & Makorova, 2004; Smith & Imura, 2004; Taguchi, 2002), which was developed originally by native English speaker teachers in the ESL context, often working at a university or a language institute. In the next section, in order to fully understand the problem and possible solutions, one needs first to consider the constraints which hinder a methodological change in the given context.

1.1.2 Constraints in implementing a communicative methodology in Japan

Constraints frequently discussed in the literature include: (a) large class sizes, (b) time constraints for teaching English communicatively, (c) a need to prepare learners for entrance examinations for higher

education, and (d) lack of Japanese teachers' ability specifically in teaching communicatively.

Large class sizes

In 2005 the average number of students per class at public lower and upper secondary schools was 30.7 and 36.7 (MEXT, 2005), and 'at the university level, classes of 120 are considered the norm' (Ryan & Makorova, 2004:51). In such large classes it is difficult for teachers to provide individuals with immediate feedback on their nontarget-like use of L2 features, such as 'intensive' *recasts* (i.e. reformulations of a learner utterance that change one or more of its components while keeping its central meaning) and *clarification requests*. These are often used in laboratory studies as a means of helping learners notice their nontarget-like questions and correct them (e.g. Mackey, 1995, 1999; Mackey & Philp, 1998; Mackey & Oliver, 2002; McDonough, 2005; Silver, 1999, 2000). Although these studies have reported the positive effects of the instruction on the learners' production of English interrogatives, the provision of intensive, immediate teacher feedback on learner error is not available in normal classrooms and there is a need for research to explore alternatives which are feasible in the classroom.

Time constraints for teaching English communicatively

Another constraint on a methodological change is lack of time. At secondary schools a teacher has to cover the authorised textbook or syllabus and it is difficult to adopt imported methods. For example, in task-based teaching which constitutes a strong version of Communicative Language Teaching (CLT), discovering the system in the process of learning how to communicate is the norm and the role of explicit teaching of grammar is discredited, on the grounds that explicitly learnt knowledge cannot be turned into the implicit knowledge necessary for automatic production (Krashen, 1982). Such discovery-based grammar learning tends to be time-consuming and it is difficult to integrate it into the current curriculum. In my teaching experience Japanese learners prefer an approach to language that is concrete and explicit, and the provision of explicit grammar teaching seems one of the effective ways to meet their need.

Since the introduction of the new communicative curriculum in Japan, what Howatt (1984) termed a

'weak communicative approach' has been practised and grammar teaching has been integrated into communicatively-oriented instruction. In a weak version of CLT, tasks are viewed as a way of providing communicative practice for L2 items that have been taught in a way proposed in notional/functional syllabuses, in which the structural properties of language are taught to learn how to realise specific general notions (e.g., past time and possibility) and functions (e.g.., apologising and requesting). The methodology employed is accuracy-based and the teaching of pragmatic aspects of the target language and of cultural/ceremonial topics are welcomed by learners and teachers, but the authorised textbooks based on notional/functional syllabuses allow teachers to spend much less time on explicit teaching of grammar, and its effectiveness has been questioned by many teachers who need to help students prepare for entrance examinations for higher education, which require them to demonstrate knowledge of grammar rather than their communicative competence in English. This is another hindrance to the required methodological change.

A need to prepare learners for entrance examinations for higher education

Teachers who give priority to helping students prepare for entrance examinations for higher education tend to teach grammar through a teacher-fronted traditional presentation-based approach: they provide learners with activities, but draw learners' attention largely to *form*, and not to *meaning*. Although these teachers recognise the importance of helping learners acquire communicative competence, improving that competence, especially teaching speaking through tasks, has been given a lesser priority, since tasks tend to be considered merely as superficial and pleasant activities - not as an effective means of facilitating a balanced development for their interlanguage systems. In fact, in order to meet students' exam entry requirements, at many upper secondary school some *Oral Communication* (OC) classes are replaced by formal grammar instruction and translation teaching from English into Japanese and vice versa. This is with the approval of the local board of education – yet some of the content of the authorised OC textbooks is left uncovered.

On the other hand, those who are eager to practise a communicative methodology have been trying to integrate grammar teaching, but the majority tend to be concerned about its effectiveness. Moreover, even among the latter type of teachers, many are concerned about outcomes in communicative tasks

which require learners to work with others of different L2 proficiency. In fact, several undesirable phenomena have been often observed in goal-oriented tasks, such as learners' imprecision in their utterances, the use of simplified language, fossilization, and the avoidance of using cognitively difficult L2 forms. This is because learners with the same L1 background can understand each other without producing complete or correct sentences (Swain, 1998; Swain & Lapkin, 1998, 2000). However, this does not necessarily mean that learning through communicative tasks is pointless for L2 acquisition; rather it suggests a need for careful lesson planning to use well-designed tasks and effective techniques to meet specific objectives.

Lack of Japanese teachers' ability in teaching English communicatively

CLT does not completely exclude use of the learners' native language (L1), but it places a methodological emphasis on intensive exposure to the target language. However, many Japanese teachers do not have native-like oral proficiency and feel unable to teach communicatively as native speakers do. They have a fear of making mistakes in using English, which has some deep socio-cultural roots[4]. Hence, instruction in L2 classes is often given in Japanese, even in L2 classes at a university-level. Furthermore, 'English is not commonly spoken by trainers or trainees in in-service teacher-training programmes given by local boards of education' (Nagasawa, 2002:287). Moreover, many of them are under-trained in implementing a communication–oriented approach because of an 'inadequate teacher training system' (Hino, 1988:52) and their knowledge of and skills in implementing a communicative methodology in their own teaching context is rather limited.

These problems suggest the need for studies of *appropriate methodology*, which has become a keyword in current ELT (Holliday, 1994) - sensitive not only to the requirements of students and teachers but also to classroom constraints in different cultures and historical backgrounds. Having identified the difficulty in implementing a communicative methodology in the Japanese context, I wish to deepen my understanding of the nature of the problems and, if possible, to find ways of addressing them which would allow effective ways of integrating grammar teaching with communicatively oriented instruction. In this study, I chose English interrogatives as targets, since opportunities to produce questions are apparently abundant in L2 classes, and learners' ability to produce questions is a basis for becoming

efficient communicators, so there is a need to explore effective ways of helping them acquire question forms (hereinafter, Q-forms). I will explain the target rules further in detail in Chapter 5 (Research Design).

1.2 Outline of the thesis

The remaining chapters of the thesis are organised as follows. In Chapter 2, I will review the literature on the theories, research and practice on which I drew. I will: (a) look at early interactionist approaches to L2 acquisition; (b) consider cognitive approaches which have been the mainstream in L2 acquisition research and have brought us better understanding of acquisitional processes; and (c) examine studies based on cognitive theory that examined effective ways of promoting noticing and modified output which is necessary for L2 development, showing how they influenced my decision in choosing a pedagogical approach for this study.

In Chapter 3, I will: (a) review research into learners' L2 development in the acquisition of English interrogatives and summarise its findings; (b) present the 6-stage developmental sequence proposed by Pienemann and Johnston (1987), which has frequently been adopted in previous studies to examine the effects of instruction on L2 development; (c) look at findings of the *Multidimensional Model* of L2 acquisition (Meisel, Clahsen, & Pienemann, 1981), on which the 6-stage sequence was based; (d) operationalise L2 development and explain my assessment criteria for individuals' developmental stage; (e) examine findings in previous studies which have informed the research design, addressing the issue of research methodology; (f) address the lack of research into the acquisition of cognitively difficult Q-forms categorised at the highest developmental stage (i.e. stage 6), such as *Q-tag* (e.g. 'There are two pubs, aren't there?') and *Cancel-Inversion* (e.g. 'Could you tell me why you haven't sent an application form to the university?'), which involve subordinate clause operations, and *Y/N-Negative* (e.g. 'Haven't you done the shopping yet?), and discuss its causes and the need for examining the effects of instruction to teach these structures.

In Chapter 4, I will review research drawing on sociocultural theory, which explores social and cultural

influences on L2 learning. The theory has been increasingly applied to L2 acquisition research and I will review the findings of such studies to find a way of providing learners with a rich learning context in the classroom.

In Chapter 5, I will describe how I set up the present study and outline the rationale for the development of the teaching and testing materials. I will: (a) present the aims of my study; (b) introduce the Q-forms targeted in the instruction and provide background information on the students who took part in the study and show an overview of timing of my study; (c) show two kinds of preliminary tests plus a questionnaire used to obtain ideas about the students' general English proficiency and knowledge of the target rules, and their background in L2 learning prior to the main study, followed by the results of their analysis; (d) explain why I chose a *noticing-promotion approach* in my instruction to teach the target rules and describe the content of four treatment sessions together with questionnaires given after each session; (e) present the tasks used to collect data in three tests given before and after the instruction in order to investigate the sustained effect of the instruction on individuals' gains; (f) discuss the questionnaires given after each test intended to explore the students' views on the value of the treatment and on their changes through it; and (g) explain how I piloted earlier versions of teaching and testing materials and what changes I made to improve them.

In Chapters 6-8, I will present results of my analysis of the effects of the noticing-promotion approach on three different areas in the acquisition of the target Q-forms. In Chapter 6, I will show the results of the examination of its effect on individuals' change in developmental stage in the acquisition of question formation. I will: (a) explain the transcribing method for the recordings of the students' performance in the three tests and the coding system of question examples; (b) explain the procedures for stage assignment for the pretest and the procedures for the measurement of individuals' sustained gain in developmental stage on the two posttests; and (c) present the results of my analysis with discussion.

In Chapter 7, I will present the results of my examination on individuals' gaps in English interrogative acquisition and on the effect of the treatment on filling those gaps, in relation to their readiness to learn the target rules. I will: (a) explain the procedures for identifying gaps in an individual's output on the pretest; (b) show how I measured the sustained effects of the treatment on filling the gaps on the

posttests and present my analysis; and (c) discuss sustained effects of the treatment on the acquisition of untargeted Q-forms in which students had shown gaps on the pretest.

In Chapter 8, I will discuss each student's perceptions of the usefulness of the instruction and of their changes in producing questions with the target rule in the classes and tests and beyond the classroom. The noticing-promotion approach required the students to take an active role both in noticing gaps between the target grammar and what they had produced, and in assisting their partner to correct mistakes while working in pairs with a peer learner most of the time. Involved as both task designer and teacher, I was concerned about my students' own views on the value of practice.

Finally, in Chapter 9, I will: (a) summarise the main research findings; (b) discuss their implications for teaching; (c) summarise limitations of the study, and (d) put forward proposals for future research.

CHAPTER 2
Current developments in research into grammar teaching

2.1 Introduction

Since Hatch (1978) proposed that L2 learning evolved out of communication, a number of studies have investigated the role of communication in L2 acquisition and the necessary conditions for meaningful interaction. With the introduction of a strong version of CLT methodology in the 1970s, which posits that acquisition takes place only through using an L2, the role of explicit grammar teaching and corrective feedback was downplayed and classrooms were filled with purely meaning-centred communication activities with no attention to language form or error correction (Howatt, 1984). This phenomenon was widespread, especially in the ESL context, while in many EFL classrooms the problem of separating out grammar teaching and communicatively oriented teaching became prevalent, since while recognising the value of a communicative methodology and being urged to practise it in classes, many teachers were not convinced that the new approach was more effective than the traditional approach to teaching L2 forms or that such an imported approach was practical in the given context.

In the 1980s, findings in early examples of immersion teaching in Canada, where content-based, purely meaning-centred teaching had been practised, showed its limitations in helping learners to attain high levels of accuracy in using some L2 features, despite years of meaningful input and opportunities for interaction (e.g. Harley, 1992; Harley & Swain, 1984). This suggested the inadequacies of such exclusive focus on meaning. Moreover, many L2 teachers found that such an exclusively meaning-centred approach was not effective in teaching new structures within limited class hours and were difficult to practise in normal L2 classes. Many researchers shifted their focus to investigating effective ways of drawing learners' attention to *form* in communication-centred instruction. Such studies have been referred to as *focus on form* (FonF) research in the L2 acquisition literature, since most researchers currently working on this issue ascribe the revival of interest in grammar teaching to the notion of FonF introduced by Long (1991). Findings in both laboratory and classroom studies have clearly shown the advantage of such instruction (e.g. Fotos, 1993, 1994; Lyster and Ranta, 1997; Lyster

and Mori, 2006; Mackey, 1999; Mackey & Philp, 1998; Mackey & Oliver, 2002; McDonough, 2005; Schmidt, 1992; Spada & Lightbown, 1993; VanPatten & Cadierno, 1993; VanPatten & Oikkenon, 1996; Swain, 1998; Swain & Lapkin, 2000; see Lightbown, 2000, for a summary of classroom SLA research).

In this chapter, in a search for effective ways of integrating grammar teaching with communication-oriented instruction for the present study, I will review FonF research and practice. I will first define *focus on form*, since the term has been used in different ways in previous studies and there has been some confusion in the literature. I will then review interactionist approaches between the early 1980s and the mid-1990s, which examined conversational features to facilitate interaction and links between those features and learner comprehension. Then I will look at a shift in L2 acquisition research from interactionist approaches to cognitive approaches, including FonF instruction, since the mid-1990s, which were based on cognitive theory that attempts to describe mechanisms and processes underlying L2 development. I will then review FonF studies in detail in search of effective ways of implementing FonF in this study.

2.2 What is 'focus on form'?

Long (1991) distinguished two types of approach to grammar teaching: *focus on forms* (FonF*S*), which characterises a traditional, synthetic forms-in-isolation type of grammar teaching in which there is little focus on meaning (i.e. pragmatic meaning, that is, the use of language in context) or communication, and *focus on form* (FonF), which 'overtly draws students' attention to linguistic elements as they arise *incidentally* in lessons whose overriding focus is on meaning or communication' (Long, 1991:45-46; italics, my emphasis). The crucial distinction between the two is that in FonF learners' engagement in meaning is a prerequisite for L2 acquisition before their attention to L2 features.

However, the term has been variously operationalised and, thereby implemented by researchers and teachers. For example, while Long defines FonF as a *reactive* teaching intervention that involves 'occasional shifts of attention to linguistic code features' (Long & Robinson, 1998:23) in response to

salient errors in using them, such as teacher feedback on learner errors, some other researchers (e.g. Swain, 1998; Swain & Lapkin, 1998; VanPatten, 1993, 1996, 2002; VanPatten & Cadierno, 1993) have included pre-selected L2 features as a target and taught them explicitly in order to raise learners' awareness of the target form. The former is referred to as *reactive* FonF and the latter as *proactive* FonF, both of which will be discussed further in detail later in Section 2.5.

Thus, there is diversity in researchers' interpretation of and approaches to FonF instruction. For the purpose of the following review study, the term Form-Focused instruction is used as an umbrella term to refer both FonF and FonF*S* approaches and I define each type of approach as follows.

> *Focus on form* (FonF) refers to pedagogical events which occur within communication-oriented instruction and in which learners' attention is drawn to form either in *proactive* or *reactive* ways.
>
> *Focus on forms* (FonF*S*), as defined by Long (1991), refers to pedagogical events which occur within a traditional, synthetic forms-in-isolation type of grammar teaching in which there is little focus on meaning (i.e. pragmatic meaning, that is, the use of language in context) or communication.

In my definition of FonF, instruction including explicit teaching of pre-selected L2 form(s), which is a major component of a traditional presentation-based FonF*S* instruction, is regarded as an option for proactive FonF as long as it is given as a part of communicatively-oriented lessons, even if it is given at the beginning of lessons, as was the case in some previous proactive FonF studies mentioned earlier (e.g. Swain, 1998; Swain & Lapkin, 1998; VanPatten, 1993, 1996, 2002; VanPatten & Cadierno, 1993). It should be noted that my definition of FonF differs from that of researchers who advocate task-based instruction (TBI), which claims the superiority of incidental (or 'reactive') FonF over proactive FonF and rejects 'proactive' FonF syllabus design and, needless to say, explicit teaching of grammar at the beginning of lessons.

The advocates of TBI claim that it is a more effective, superior approach than 'traditional' methods because it is soundly based in theory and research, but several researchers have questioned its validity. For example, Swan (2005:376) has pointed out that "the claim is based on unproved hypotheses, and there is no compelling empirical evidence for the validity of the model. Many advocates of TBI reject proactive syllabus design on doctrinaire grounds, while commonly misrepresenting 'traditional'

classroom practice". He has also pointed out limitations of TBI as follows: although TBI may be effective to develop learners' command of what is known, 'it is considerably less effective for the systematic teaching of new language' (ibid. 376).

Sheen (2005) has also pointed out the absence of published empirical evidence that demonstrates the superiority of TBI (or 'reactive' FonF approaches) to 'proactive' approaches, including FonF*S* ones. In fact, there has been no empirical study comparing the effects of a FonF approach with those of a FonF*S* approach. Moreover, findings in Norris and Ortega's (2000) meta-analysis of many relevant studies have shown that 'a focus on form and focus on formS are equally effective' (ibid.:501) and that 'the current state of empirical findings indicates that explicit instruction is more effective than implicit instruction' (ibid.). Sheen (2005:300) has criticised the lack of studies to put the advocacy of new teaching strategies under critical scrutiny, pointing out that the applied linguistic literature of the day 'has accepted largely without question the premature rejection of previously-used teaching options'. He has also warned that it is not appropriate to advocate reforms which exclude a FonF*S* approach from the range of teaching strategies available to the teacher, as is the case in TBI. Both Swan (2005) and Sheen (2005) have called for further research on the effectiveness of form-focused instruction which compare effects of a FonF approach with those of a FonF*S* approach.

Before examining various ways of implementing proactive and reactive FonF instruction and theories on which FonF research is based, I will review in the following section early attempts to examine the role of interaction in L2 acquisition.

2.3 Interactionist approaches

2.3.1 Early studies

In the early 1980s, there was increasing support for a theoretical position in L2 acquisition research in North America which maintained that the necessary and sufficient condition for L2 acquisition was exposure to *comprehensible input* that contains slightly more advanced grammatical forms than the

learner's current state of interlanguage, 'i+1' in Krashen's (1985) terms, which is known as the *Input Hypothesis*. Researchers who took this position argued that the provision of metalinguistic explanation of L2 rules and corrective feedback on learners' errors was ineffective, or even detrimental, since it could interfere with the process of their natural interlanguage development (Krashen, 1981, 1982; Krashen & Terrell, 1983). These claims were based on findings in English morpheme acquisition studies (Bailey, Madden & Krashen, 1974; Dulay & Burt, 1975) and in syntactic acquisition studies (Felix, 1981; Schumann, 1979; Wode, 1976), which found a consistent order of acquisition, regardless of learners' L1 background and age. Acquisition orders found in these studies were similar to those observed in L1 acquisition, and researchers claimed that an L2 could be best instructed in the way natural L1 acquisition took place.

This position had an enormous impact on L2 pedagogy. In many L2 classrooms, especially in the ESL (English as a second language) context, teachers stopped both teaching grammar explicitly and giving corrective feedback on learner errors in using L2 features. Instead, they engaged learners in communication activities designed to facilitate active interaction in the target language between a teacher and learners, and between learners, in order to provide comprehensible input, as embodied in the strong version of CLT. Such input-driven approaches assume that L2 development will take place under the right conditions by encouraging the learner to engage in comprehension through exposure to meaningful input, as is the case in L1 acquisition. On the other hand, many teachers in the EFL (English as a foreign language) context were sceptical about the effectiveness of the new imported method in their classrooms and kept practising traditional teaching while trying to integrate communicative activities into their teaching.

2.3.2 Negotiation of meaning studies

Being strongly influenced by Hatch's (1978) work on the importance of conversational interaction and the *Input Hypothesis*, Long (1983, 1985) proposed the *Interaction Hypothesis*. It claims that interactional modifications, which Long terms *negotiation of meaning* to refer to the ways in which interlocutors modify their utterance if it is not understood, play a facilitative role in L2 acquisition. This is because they increase the amount of *comprehensible input* through learners' attempts to make

comprehensible what was initially incomprehensible to the interlocutor. Many studies conducted in the 1980s and early 1990s drew on the Interaction Hypothesis and focused on the examination of the links between negotiation of meaning and comprehension.

Early studies examined conversations between native speakers (NS) and nonnative speakers (NNS) and described how learners at lower levels differed from those at higher levels (e.g. Doughty and Pica, 1986; Gass & Varonis, 1985; Long, 1980; Pica, 1987, 1988; Pica, Young, & Doughty, 1987; Varonis & Gass, 1985). The researchers analysed patterns of negotiation routines in conversations, their frequency and functions, and identified conversational features to facilitate negotiation of meaning. These include *confirmation checks* (made by the speaker in order to confirm whether or not their understanding of the interlocutor's utterance is correct), *comprehension checks* (made by the speaker to check whether the interlocutor has understood their preceding utterance), *clarification requests* (made by the listener for further information from an interlocutor regarding a previous utterance), requests for *repetitions*, and signals to show *non-understanding*.

For example, Varonis and Gass (1985) examined the amount of negotiation of meaning in three conditions: between NSs, between NS and NNS, and between NNSs. It was found that little negotiation of meaning took place in NS-NS discourse, while the greatest number of instances were observed in NNS pairs. It was also reported that negotiation of meaning was more common in NNS-NNS interaction than in NS-NNS interaction, especially in paired conversation where participants were from different L1 backgrounds or did not have the same proficiency in English.

Another descriptive study by Doughty and Pica (1986) examined effects of task type and participation patterns on L2 classroom interaction. They counted the number of comprehension checks, clarification requests, confirmation checks and repetitions that initiated negotiation of meaning in different tasks in three interaction conditions: teacher-fronted work, group work and learner pairs. They compared the results of this study to those of their early study (Pica & Doughty, 1985) in regard to optional and required information exchange across the three conditions and found that a task which required information exchange (two-way information gap task) resulted in more modified interaction than other types of task without such a requirement. These findings suggest that both participation patterns and

task type have an effect on the production of modified interaction.

Gass and Varonis (1994) examined whether conversational interactions between NS and NNS in a two-way information gap task would lead to better NNS comprehension and to more accurate English production. In this task, which was repeated twice, learners and the NS were each provided with identical picture boards depicting an outdoor scene. The task consisted in placing 20 object cut-outs on the board, following a script read by the NS describing the location of the objects on his/her board. Some learners were allowed to interact with the NS during the description (to use requests for repetition and clarification, comprehension checks, etc.), while others relied solely on the scripted description. The NS-NNS dyads then changed roles and the NNS described a similar scene for the NS to place the object cut-outs. It was found that the learners allowed to interact outperformed those who were not, suggesting better comprehension. However, in a second trial, it was also found that those learners who had been allowed to interact in the first trial did not necessarily produce more accurate English in describing unknown lexical items.

2.3.3 Criticism of interactionist approaches

Thus, interactionist studies in the 1980s and early 1990s contributed to a clearer understanding that modified interaction would serve as comprehensible input for better comprehension and that opportunities for *comprehensible output* are necessary for L2. However, they have been criticised for lack of evidence that much modified interaction or negotiation of meaning leads to long-lasting gains in learning specific L2 items, that is, L2 development. For example, a review of research on the effects of tasks intended to facilitate negotiation of meaning by Pica, Kanagy and Falodun (1993) found no evidence linking *negotiation of meaning* in task interaction with L2 development. They concluded that '[D]espite the important contributions that task-based research has made in supplying data and supporting theories on second language acquisition, few studies have actually linked negotiation features found during task interaction with acquisition processes themselves' (27).

In addition, many researchers who examined the effects of task type and interaction conditions on learners' production of negotiation of meaning in classrooms expressed concerns about the effectiveness

of interaction-based approaches to L2 teaching (e.g. Gass, Mackey and Pica, 1998; Gass & Varonis, 1994; Harley & Swain, 1984; Loschky, 1994; Sheen, 1994; Swain, 1984, 1985; Swain & Lapkin, 1995, 1998, 2000; White, 1987).

For example, Foster (1998) observed classroom interaction generated through tasks by 21 ESL learners with different L1 backgrounds and found that 'many learners in small groups did not speak at all, [and] many more in both dyads and small groups did not initiate any negotiated interaction, and very few students in either setting produced any modified utterances' (1). She has offered several reasons for the lack of learner interaction in her observation. One is that some students dominated the conversation during the task by bombarding the other student with questions, requiring the partner only to give '*yes/no*' replies. Another reason suggested is the students' unwillingness to engage in 'the difficult and potentially frustrating task of modifying their language in order to make it comprehensible' (ibid. 17). Foster also suggests that the lack of modified interaction was the students' indifference towards other learners' utterances, and that there was a possibility that students may have felt annoyed at having to give their attention to each communication breakdown. Given the big difference between what she observed in the classroom and what was reported in the experimental studies, Foster concluded that, 'contrary to much SLA theorizing, … 'negotiation of meaning' is not a strategy that language learners are predisposed to employ when they encounter gaps in their understanding' (ibid. 1).

A lack of negotiated interaction in L2 conversation has also been reported in other classroom studies (Doughty & Pica, 1986; Kowal & Swain, 1994). Doughty and Pica (1986) observed very few instances of interactional modification in whole class and group discussions in six ESL classes with intermediate students, although more modified output was found in discussion in dyads. They concluded that whole class and group activities do not automatically result in the modification of interaction among the participants. They also found that the learners' production of conversational modification was much less across three conditions in the task which *did not* require an exchange of information than one which *did*. This suggests a need to design tasks which demand information exchange.

Another concern about interactionist approaches was raised by White (1987). She maintains that some

L2 structures may never appear in the input and such forms might not be accurately acquired if the learner is unaware of the absence of certain structures in the input. White also points out that, if the input is too far advanced compared to their interlanguage, it may not be noticed or be understood. These are serious flaws in approaches driven by input and interaction where no explicit grammar instruction is provided. The issue of learners' readiness will be further discussed in the next chapter.

To sum up, early interactionist research was unable to demonstrate strong evidence that interaction is beneficial for L2 development and this criticism has led researchers, instead of looking for evidence of negotiation of meaning, to find evidence that interaction can actually be beneficial for sustained gains in the acquisition of specific L2 features in L2 development.

2.3.4 Recent developments in interaction research

In the mid-1990s researchers began to demonstrate empirically that interaction was beneficial for L2 production and development. A number of studies were conducted in laboratory settings as well as in L2 classrooms. One characteristic of recent interaction research is that it has been informed by cognitive psychology and other related fields, taking account of underlying cognitive processes arising from the interaction. This new movement, in which researchers attempt to explain various aspects of L2 learning through exposure to the target language, through production of an L2, and through feedback on learner output, is well embodied in Long's (1996) updated *Interaction Hypothesis* and Swain's (1985, 1995, 1998) suggestion as to the role of *comprehensible output* in L2 acquisition.

Long states in his updated *Interaction Hypothesis* that:

> '*negotiation for meaning*, and especially negotiation work that triggers interactional adjustments by the NS or more competent interlocutor, facilitates acquisition because it connects input, internal learner capacities, particularly selective attention, and output in productive ways' (ibid. 421-452).

Swain (1985) proposed the *Comprehensible Output Hypothesis*, which claims that learners need to be 'pushed' in their efforts to be understood by interlocutors so that they notice the inadequacy of their own interlanguage system. It is important to recognise that Swain's claim is that production aids L2

acquisition *only when the learner is 'pushed'* to make their output more accurate and coherent, not when they are merely provided with ample opportunities for negotiation of meaning. Learners can comprehend input without analysing syntactic features, but making one's own output comprehensible to others requires the learner to reflect on their output and this 'may be the trigger that forces learners to pay attention to the means of expression' (Swain, 1985:249).

This hypothesis is based on Swain's findings of the limited success of aforementioned early French immersion programmes. In the French immersion classes, the school children were taught subjects through L2 French and an emphasis was placed on meaning over form, and comprehension over accurate production; the students were not provided with any explicit teaching of L2 grammar and corrective feedback on their L2 use. Studies by Swain and her colleagues (Harley & Swain, 1984; Swain, 1984) reported that, despite abundant exposure to meaningful input, many of the students in the immersion classrooms failed to attain native-like proficiency in either oral or written French, especially, in accuracy in using some morphological and syntactic features. These findings led Swain and her colleagues to suggest that an input-rich, interaction-abundant, communication-centred classroom was not sufficient for the development of target-like proficiency.

In more recent articles (1995, 1998), Swain has proposed three functions of *comprehensible output* which may serve the language learning process: *noticing*, *hypothesis formulation and testing*, and *metatalk* (conscious reflection). The first function is *noticing*. She hypothesises that output promotes noticing the gap between what learners want to say and what they can actually say in the target language. This may lead them to notice what they do not know precisely how to say or write the message they want to convey, which she termed 'holes' (Swain, 1998:66) in one's interlanguage. The second function of output is *hypothesis formulation and testing*. Through output, learners may try out new L2 forms (hypotheses) and this may attract feedback that can lead them to reprocess their output. The third way in which output may serve the acquisition process is through *metatalk*. As learners reflect upon their own L2 use, or their conversation partner's, by using metalinguistic terminology, it may enable 'them to control and internalise linguistic knowledge' (Swain, 1995:126). Swain (1998) considers that 'metatalk is a surfacing of language used in problem solving, that is, it is language used for cognitive purposes' (69); here she is drawing on the sociocultural theory of mental activity, which will be

discussed in further detail later in Chapter 4.

As we can see, Long (1996) and Swain (1985, 1995, 1998) attribute considerable importance to taking account of learning processes in L2 acquisition research. In the next section, before examining findings in recent interaction research, in order to deepen my understanding of cognitive processes underlying L2 acquisition, I will review theories in cognitive psychology on which recent interaction research has substantially drawn.

2.4 Cognitive approaches

An advantage of cognitive approaches to L2 acquisition is that they are supported by research findings in well-established cognitive learning theories, which allow us to understand what factors contribute to learners' internal mental processing involved in L2 acquisition. These approaches differ from other linguistic approaches, since the latter are not concerned with investigating the inner workings of the brain as the learner engages in interaction. Whereas input-driven instructional approaches and negotiation of meaning approaches see L2 acquisition as 'the product of engaging in meaning-processing – in the first case through comprehension, and in the second through production' (Skehan, 1998:12), cognitive approaches are concerned with cognitive processes which occur in 'three stages of information processing: input, central processing involving working and long-term memory operation, and output, as well as the interaction between these stages, and the way material is presented in memory' (Skehan, 1998:44).

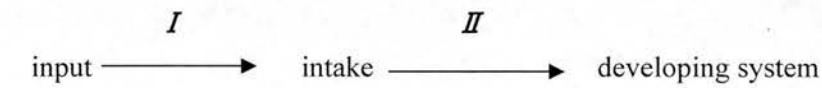
In the first two types of approach, already discussed, comprehensible input provided at the right stage of development is seen as a catalyst to trigger the process of acquisition, as well as a driving force to bring about interlanguage change. This early view of the role of input did not provide an explanation as to how such change occurs and what the 'right time' is, and it is assumed that learning is an unconscious process. On the other hand, in learning models based on cognitive theory, the process of interlanguage change and development is explained in terms of constraints on the human information processing system, such as learners' limited memory and mental resources to allocate their attention to an L2 feature

at each stage of learning.

2.4.1 VanPatten's (1996) input processing model

One well-known learning model is VanPatten's (1996) *input processing model*, which is shown below.

FIGURE 2.1 VanPatten's (1996) input processing model



VanPatten claims that by implementing a pedagogical intervention which focuses on making the form more salient at the *input*-to-*intake* stage (stage *I*), what has been processed from the input can more easily and efficiently become *intake*, and that the learner's explicit attempts to incorporate what they have attended to at stage *I* into a developing interlanguage system in stage *II* (i.e., *intake*-to-*restructuring* stage) can bring about L2 acquisition. This model suggests the usefulness of instruction which engages the learner in noticing form-meaning links included in the input.

2.4.2 Noticing and interlanguage restructuring

In a similar vein, Schmidt (1990) proposed a crucial role for *noticing* in the process of L2 acquisition. He claims that 'noticing is the necessary and sufficient condition for the conversion of input to intake for learning' (Schmidt, 1994a:17), now well known as the *Noticing Hypothesis*.

He has further maintained that it is indispensable to draw the learner's *attention* to linguistic mismatches between their interlanguage and the target language, the process known as *noticing the gap*, since not all input has equal value and only what is noticed in the input becomes available for *intake* and effective processing prior to the incorporation of the element into developing interlanguage systems. For instance, Schmidt has reported the necessity of paying attention to a language feature for the adult

acquisition of redundant grammatical features, such as the third-person singular morpheme '*-s*' (Schmidt, 1992).

Doughty and Williams (1998) also argue that reconstructing 'essentially involves the reorganization of those developing IL rules to be most effectively deployed during language use and, eventually, to become more targetlike' (228) and suggest the following four subprocesses which are engaged during the learner's analysis of systematic rules and are 'catalysts for restructuring' (228): (a) hypothesis testing, (b) cognitive comparison, (c) noticing IL/TL differences (often called *gaps*), and (c) noticing IL deficiencies, or 'the hole' (Swain, 1998:66). Doughty and Williams (1998) emphasise that these processes are vital for helping learners gain 'control over knowledge in order to use it fluently and skilfully' (229), that is, for L2 acquisition.

Now a question arises: are these acquisitional processes leading to interlanguage *restructuring* conscious or subconscious? This is a key issue in research taking a psycholinguistic or cognitive orientation.

2.4.3 The role of consciousness and attention

Schmidt (1990) points out that the role of unconscious learning has been overstated, as seen in Krashen (1985), and argues that acquisition is largely a conscious process, distinguishing four senses of *consciousness* with regard to L2 acquisition (Schmidt, 1994a):

(a) *consciousness* as intentionality (i.e. the learner can intentionally set out to learn something or the learner can learn something incidentally while focusing their *attention* on some other learning goals);
(b) *consciousness* as attention (i.e. L2 acquisition involves conscious *attention* to L2 form, irrespective of whether it is intentional or incidental);
(c) *consciousness* as awareness (i.e. the learner is aware of what they are learning); and
(d) *consciousness* as control (i.e. the use of knowledge in L2 production involves conscious processes of selection and assembly).

Schmidt (1990) suggests that the degree of learners' *focal* attention and awareness plays a key role in the acquisition processes, although he recognises that not all learning is conscious. For instance, intentions can be unconscious and learners often become aware of L2 items they do not intend to notice.

Tomlin and Villa (1994) have objected to the above proposal that acquisition is largely a conscious process and argued that conscious registration is not 'always' a prerequisite for attentional processes in L2 acquisition and that it is in part conscious and in part subconscious. They have pointed out that early conceptions of attention were coarse-grained, since the discussions were focused on the limited-capacity metaphor or the *automatic* versus *controlled processing* dichotomy. They elaborated the mechanisms of *attention* and divided attention into three components which constitute separate and yet interrelated networks:

(a) *alertness* (i.e. the learner's affective/motivational readiness to deal with incoming stimuli or data),
(b) *orientation* (i.e. aligning of attention on some specific type or class of sensory information at the expense of other), and
(c) *detection* (i.e. the cognitive registration of sensory stimuli in short-term memory).

Tomlin and Villa (1994) have argued that all three processes, including detection where noticing occurs, can take place without awareness, commenting that awareness requires attention, but attention does not require awareness.

McLaughlin (1990) has maintained that the distinction between *consciousness* and *unconsciousness* is not helpful in research in examining acquisitional processes and that it is more appropriate to consider L2 acquisition in terms of whether it involves *controlled processing*, which can often require *attention*, or *automatic processing*, which does not require *attention*, since 'both controlled and automatic processes can in principle be either conscious or not' (620). McLaughlin and Heredia (1996) also argue that more *attention* is required in some tasks than others which have been well-practised, and that it is necessary to help the learner develop complex cognitive skill that involves building up a set of well-learned, efficient procedures which allow them to free up more attention-demanding processes for new tasks. The explanation that controlled processes become automatic through practice accords with Anderson's (1993, 1995) widely accepted *Skill Acquisition Theory*, which will be discussed later.

Thus, researchers' opinions vary as to the role of consciousness in L2 acquisition, but they agree on the need for empirical examination for further understanding of acquisitional processes. For example, McLaughlin (1990) proposed that research into the following four areas will be useful: whether learning

is implicit or explicit, whether learning is intentional or incidental, whether the learner is aware or unaware of learning, or whether learning concerns the short or long term memory. It is beyond the scope of this study to review all these areas in detail, but in the following two sections I will review research on *implicit* and *explicit learning* and on *implicit* and *explicit knowledge*, which have been key issues in L2 acquisition research, making reference to the other issues mentioned above. After that, I will summarise the pedagogical implications arising from the review.

2.4.4 Implicit and explicit learning

It is generally accepted that there are two types of learning: *implicit* and *explicit learning*. Implicit learning is considered 'a nonconscious and automatic abstraction of the structural nature of the material arrived at from experience of instances' (Ellis, N., 1994:2-3), which is incidental, and therefore, it 'does not involve selective attention to features of input that feed into the learning process [...] and is unaffected by instruction' (Schmidt, 1994b:172). Explicit learning is considered 'conscious searching, building then testing of hypotheses, assimilating a rule following explicit instruction' (Ellis, N., 1994:2-3), and therefore, it does involve selective attention to rules.

Implicit learning is advocated in non-interventionist approaches, such as the strong version of CLT which is strongly influenced by Krashen's (1985) claim that only implicitly learnt knowledge can be converted into *acquired* knowledge, which can be easily retrieved in spontaneous communication. On the other hand, explicit learning is represented in 'the wide range of teaching methodologies which rely on rule presentation either before practice and production activities, or after communicative exploration, at which time they may have consciousness-raising and consolidating roles' (Skehan, 1998:54-5). A number of experimental studies have been carried out to examine the relative effectiveness of these two types of learning.

In the review of such studies by Long and Robinson (1998), it is suggested that explicit learning through focus on L2 features is more effective than implicit learning for simple rules which have clear form-function relationships, such as plural *–s*. For complex ones, research findings are mixed. For

instance, in an experimental study on the acquisition of English dative alteration, which is considered a complex rule, Carroll and Swain (1993) reported an advantage of explicit teaching in combination with explicit metalinguistic feedback over instruction consisting of explicit teaching in combination with implicit feedback, such as *recasts*. Ellis (1993) also found an advantage of explicit learning (via explicitly being taught a complex rule of Welsh, initial consonant mutation, plus carefully considered examples of rules) over *implicit learning* (via exposure to thousands of examples of the rule), but in other studies (Dekeyser, 1995; Robinson, 1996) learning difficulty was observed in both explicit and implicit learning conditions.

Thus, although the research findings were not clear-cut as to the acquisition of complex forms, it is possible to say that these findings suggest an advantage for explicit over implicit learning. However, we should be cautious about generalization from such experimental results to classroom L2 learning - a point critically made by VanPatten (1994). For example, in these studies conditions were tightly controlled and it was useful to examine the effects of techniques used in each condition for comparison, but such laboratory conditions seldom exist in normal L2 classes. Moreover, the means used for the assessment (e.g. grammaticality judgment test in Ellis, N., 1994 and in Robinson, 1996)[5] did not resemble everyday language use. To sum up, it is realistic to think that it is not possible for the learner to learn a complete language explicitly or intentionally, and therefore, it is reasonable to think that learning which takes place implicitly or incidentally also has an important role in L2 acquisition.

2.4.5 Implicit and explicit knowledge

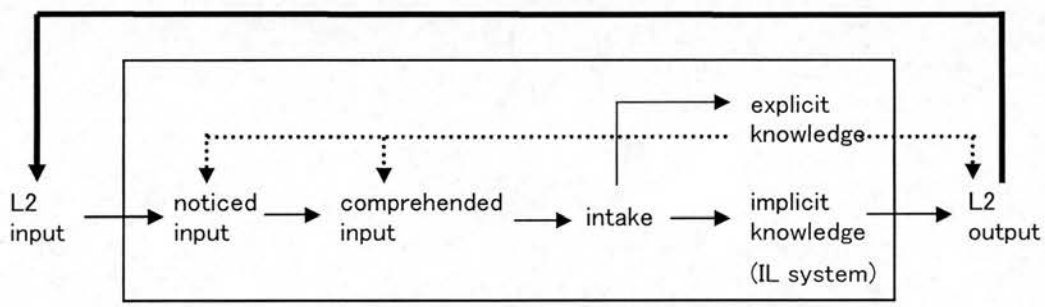
The other important issue in cognitive approaches to L2 teaching is the role of *implicit* and *explicit knowledge*. In L2 acquisition research, the existence of these two types of knowledge has been generally accepted (e.g. Anderson, 1987; Bialystok, 1979, 1982, 1991, 1994; Krashen, 1982; McLaughlin, 1978, 1987).

Explicit knowledge is "generally used to refer to knowledge that is available to the learner as a conscious representation. It is not the same as 'metalinguistic knowledge' (knowledge of the special terminology for labelling linguistic concepts), although it is often developed hand in hand with such knowledge"

(Ellis, R., 1994:355). Implicit knowledge comprises two types of knowledge: 'formulaic knowledge and rule-based knowledge. The former consists of ready-made chunks of language … Rule-based implicit knowledge consists of generalised and abstract structures which have been internalised. In both cases [i.e., the two types of implicit knowledge], the knowledge is intuitive and, therefore, largely hidden; learners are not conscious of what they know. It becomes manifest only in actual performance' (Ellis, R., 1994:355-6).

Based on Gass (1988), Ellis, R. (1994) proposed a theoretical framework for investigating L2 acquisition, which describes how explicit and implicit knowledge is involved in processing input in L2 acquisition, as shown below.

FIGURE 2.2 A framework for investigating L2 acquisition (Ellis, R.1994:349)



Four stages of information processing are distinguished by Gass (1988): (1) *apperceived* (or *noticed*) *input*, (2) *comprehended input*, (3) *intake*, and (4) *integration*. Only noticed features in the input in the first stage are comprehended and '[N]ot all apperceived input is comprehended (or contributes to the learner's understanding of message content)' (Ellis, R. 1994:349), but not everything comprehended becomes intake. However, 'what is comprehended can either feed into the intake component or, alternatively, it may be not used by the learner for anything beyond communication' (Gass, 1988:205). The subsequent processes are described by Ellis, R. (1994:349) as in the following:

"Intake, following Chaudron (1985), is seen as 'a process which mediates between target

language input and the learner's internalised set of rules' (Gass,1988:206). It does not become part of the learner's implicit knowledge system until it has been 'integrated'."

One additional component proposed in the model is explicit knowledge. According to Gass (1988), some input which has been processed but not integrated into the interlanguage system is put into 'storage', which 'takes the form of some kind of explicit representation of L2 features. Explicit knowledge 'can contribute to output through monitoring, and also may aid the processes that contribute to intake. Output can influence input through interaction' (Ellis, R. 1994:349). In short, this model suggests the important role of both explicit and implicit knowledge through acquisitional processes.

As for the representation of explicit and implicit knowledge, however, researchers take different theoretical positions and address different aspects of this framework regarding their roles in L2 acquisition. There are three theoretical positions: the *non-interface*, *interface*, and *variability* positions.

Krashen holds the strongest non-interface position, claiming that learned (i.e. explicit) knowledge and acquired (i.e. implicit) knowledge are totally separate and that explicit knowledge cannot be converted into implicit knowledge, which is based on his Monitor Theory of SLA (Krashen, 1982). Krashen further asserts that drawing learners' attention to L2 form has little point in L2 acquisition since acquired knowledge can only be developed when the learner's attention is focused on meaning. In other words, the non-interface position assumes a very limited role for providing learners with explicit knowledge through grammar teaching.

The interface position holds that explicit knowledge can be converted to implicit knowledge and vice-versa (e.g., Anderson, 1987; Bialystok, 1979, 1982, 1991, 1994; DeKeyser, 1998; Ellis, 2003; McLaughlin, 1978, 1987; Sharwood Smith, 1981). For example, McLaughlin (1978, 1987) proposes a model of L2 acquisition which describes the process of acquisition as a gradual transition from *controlled processing* to *automatic processing*, based on information processing theories. Bialystok (1994) also suggests that both implicit and subsequently acquired language knowledge could become more explicitly represented in terms of the target language through the two psycholinguistic processes of *analysis* and *restructuring*. For instance, by analysing the parts that make up formulas, learners can

convert unanalysed knowledge into analysed knowledge.

In a similar vein, Anderson (1987) proposes that implicit (*procedural* in his terminology) and explicit dimensions are not completely independent, based on his widely accepted *Skill Acquisition Theory* (Anderson, 1993, 1995), which claims that skill acquisition consists of three stages: the attainment of *declarative knowledge* (i.e. knowledge about language and the use of language), *proceduralization of knowledge* (i.e. ability to perform declarative knowledge), and *automatizing or fine-tuning procedural knowledge* (i.e. ability to use such knowledge without having to think about it). Anderson (1987) explains that *proceduralization* can be achieved by engaging in the target behaviour of using language while temporarily leaning on declarative crutches. *Automatization* of the newly acquired proceduralized knowledge, which is necessary for interlanguage restructuring, can then be realised, depending on the amount of practice and the necessity of using that knowledge, increasing speed and decreasing the error rate and the demand on cognitive resources. These claims lead to the logical conclusion that the teacher must enable the learner to engage in the practice of using target language while providing them with declarative knowledge to be kept in working memory.

While these researchers adopt a strong interface position (e.g., DeKeyser, 1998; Sharwood Smith, 1981), Ellis (2003) has proposed 'a weak-interface position, according to which explicit knowledge facilitates the development of implicit knowledge rather than changes into it .. [and] serves to prime attention to form in the input and thereby to activate the processes involved in the acquisition of implicit knowledge' (ibid., 106).

The variability position considers that the way an L2 is learnt is a reflection of the way it is used. This position acknowledges that each developmental stage involves both *systematic* (or contextual) variability and *non-systematic* variability, both of which consist of 'the rearrangement of a previous variable system into a new variable system' (Ellis, 1985: 97). This rearrangement can take place in two ways. The rearrangement of systematic variability can occur in a way that the learner's dominant use of one style (e.g. the careful type) will be gradually taken over by another type (e.g. the vernacular type) depending on the learner's specific needs and preferences in certain social contexts and/or their capacity to pay *attention* to L2 forms in their speech. Non-systematic variability, on the other hand, will be

learnt while learners engage in the reshuffling of form-function relationships by eliminating a redundant form when acquiring a new one, in order to achieve the maximum effect on the development of interlanguage systems, allowing non-systematic variability to become systematic variability. Thus, the variability position is quite different from the other positions, since it takes into consideration the effects of sociolinguistic factors on the acquisition of fuller L2 knowledge.

2.4.6 Delayed acquisition

Before summarising pedagogical implications, it would be relevant here to cite Kellerman's (1985) description of *U-shaped learning behaviour,* observed in L2 structure learning in the form of a tripartite sequence. The *u-shape* refers to the shape of a graph in which accuracy is measured against time. At the early stage of this sequence, learners may manifest correct use of a target-language structure or lexical item, which exhibits a level of grammatical accuracy that is clearly beyond the learner's proficiency as evidenced by his output in toto'(ibid.: 347). In the second stage, however, this accuracy can be lost for a while in some cases, which represents the downward turn in accuracy in development. Kellerman explains that the accuracy in the first stage may be due to chunk-learning and that the regression in the second stage may be influenced by more sophisticated attention to L2 forms, especially to the distinction between form and function. In the third stage, learners' accuracy increases gradually as they become able to intake L2 features and their interlanguage reconstructing takes place. Kellerman's discussion suggests a need for careful examination of learner output when judging whether a certain structure has been acquired or not. In order to capture delayed acquisition of L2 features, it is important to have more than one posttest (i.e., an immediate posttest and follow-up tests) to examine effects of instruction.

2.4.7 Studies on acquisition order of L2 forms

It is also useful to extend my discussion to findings in research into the acquisition of L2 forms. Developments in theoretical linguistics, especially in theories on language universals, have also increased our understanding of how a language is learned. For instance, Zobl (1983) introduced the notion of 'projection', which accounts for the disparity between the input data available to learners and

the ultimate state of knowledge attainment. Then, Zobl (1985) proposed the *Markedness Theory*, which stated that the limitations of the input data could be overcome by learners through two types of acquisition procedures: (a) 'acquisition via markedness implications and (b) acquisition via correlations between markedness of different but related parameters' (Zobl, 1985:330). In other words, the provision of markedness conditions in input data may 'provide the crucial and necessary shortcuts which make possible the ultimate states of knowledge' (Zobl, 1985: 344).

In fact, research into *relative clause* acquisition has shown that form-focused instruction is effective when it targets stages which are 'several steps beyond' the learners' current level of development (Doughty, 1991; Eckman, Bell and Nelson, 1988; Gass, 1982;). Most of these studies were based on Keenan and Comrie's (1977) *Noun Phrase Accessibility Hierarchy* (NPAH). They have found that relative clauses in certain positions are more marked than those in other positions (see Figure 2.3 below).

FIGURE 2.3: The Noun Phrase Accessibility Hierarchy (NPAH)

Subject > Direct Object > Indirect Object > Object of Preposition > Genitive > Object of Comparison

More accessible (Less marked)

Less accessible (More marked)

The hierarchy is to be interpreted in such a way that if a language allows relativisation on a certain (more marked) position, it also allows relativisation on certain (less marked) positions to the left of this more marked position. For instance, the subject clause is the least marked and is the most accessible in its hierarchy. In most grammar-based textbooks, only subject relatives are dealt and, if there are any, direct object relatives are presented before genitive and object of preposition relatives. Object of comparison relatives are rarely dealt with in textbooks. 'Underlying this ordering is the implicit assumption that by teaching the easy ones first, appropriate generalizations will be made to the more difficult relative clause positions, those which are either not taught at all or which are given the least emphasis' (Gass, 1982:131).

Using the NPAH framework, Gass (1982) studied the instructional effects on a hypothesized generalisability, or projectability, of one type of relative clause to another. Two groups of L2 learners were given instruction focusing on different relative clauses. One group was taught subject and direct object relatives, which were supposed to be less difficult to learn and the other group was taught object of preposition relatives, which were thought to be most difficult. It was found that the group receiving the instruction on less difficult relative clauses (more accessible, less marked construction) performed well on the target forms, while the group that received the instruction on the more difficult relative clause construction (i.e., less accessible, marked construction) performed well not only the target forms, but also on the other types of relative clauses as well. These results suggest that learners can gain greater control over an L2 structure in relative clause acquisition when only the more difficult aspects are taught. can be of instruction showed that the opposite occurred in English relative clause acquisition. Similar results were reported in studies by Eckman, Bell, and Nelson (1988) and Doughty (1991). One pedagogical implication drawn from the findings is that the markedness found in a pair/set of language forms can be effectively used when designing tasks. It is possible to infer that teaching a more marked form with consideration to learners' readiness (i.e., forms which are supposed to be not only at the next level in the hierarchy, but those beyond the next level) helped learners to acquire the less marked form.

On the other hand, when it comes to the acquisition of English interrogatives, research findings have shown different results. For example; Pienemann (1984, 1988, 1989) and Spada and Lightbown, (1993) found that instruction was effective only when it targets forms at the next stage of learners' L2 development. Findings in these studies are clearly the most relevant to my study and I will examine them further in detail in Chapter 3 together with relevant theories concerning interrogative acquisition.

2.4.8 Pedagogical implications

We have seen that instructional effects on the acquisition of L2 forms differ between relative clause acquisition and interrogative acquisition, and that researchers' opinions differ considerably as to the representation and role of explicit and implicit knowledge in L2 acquisition, but the following pedagogical implications emerge from the review.

L2 instruction should:

(a) provide the learner with conditions for both explicit and implicit learning;
(b) provide comprehensible input;
(c) find means to promote noticing in order to facilitate intake;
(d) provide opportunities for automatizing controlled processes;
(e) help the learner develop explicit knowledge and encourage the learner to make use of it for L2 development;
(f) take into account learners' readiness to learn specific forms.

Having obtained a better understanding of acquisitional processes in L2 learning, I will now turn to examine findings in FonF research.

2.5 Focus on form instruction and SLA

Inspired by the cognitive theories and their implications for L2 instruction, a number of studies have sought effective ways of directing the learner's *attention* to grammatical features in communicatively oriented classrooms in order to promote their interlanguage development. These studies are often referred as *focus on form* (FonF) research, as mentioned in Section 2.2. In previous studies, various tasks and techniques have been used to draw the learner's attention to *form* at different levels of explicitness. In order to deepen my understanding of effective ways of implementing FonF instruction, I will now examine the advantages and disadvantages of various tasks and techniques used in previous studies.

FonF approaches can be categorised as *reactive* and *proactive*.

2.5.1 Reactive focus on form

In reactive FonF, learners' attention is drawn to form, often 'on-the-spot', in reaction to learner errors in a lesson. It is a *responsive* teaching intervention, which Long (1991) seems to have had in mind in his definition of FonF. Most examined reactive FonF techniques are various types of teacher feedback on learner error in communication. One strong point of *reactive* FonF is that learners' attention is drawn to linguistic features as they arise *incidentally* in the course of communicating a message without

interrupting the flow of communication, which is in accordance with the principles of CLT. However, some forms of reactive teacher feedback techniques, such as recasts, are implicit and learners may not perceive them as a signal that what they have said was nontarget-like because they may not notice what has been reformulated by the teacher in recasts.

A range of question techniques and types of teacher feedback on learner error have been identified in observational studies (Lapkin & Swain, 1996; Lyster, 1994, 1998b, 1998c). Those *reactive* FonF techniques are categorised into three by Lyster and Mori (2006): (a) *explicit correction*, (3) *recasts*, and (c) *prompts*.

Explicit correction and recasts provide the learner with a reformulation of their nontarget-like utterances. Whereas in explicit correction learner error is corrected by the teacher with a clear indication that what a learner has said was incorrect, in *recasts* all or part of the learner's erroneous utterance is implicitly reformulated by the teacher (e.g. Learner: 'In your picture children playing?' —Teacher: 'Are the children playing?'). As for prompts, Lyster and Mori (2006) list four techniques: (a) *clarification requests* (e.g. 'Pardon?; 'I don't understand'), which indicate to the learner that their utterance included errors in some way and that reformulation is necessary, (b) *repetition*, in which the teacher repeats the learner's erroneous utterance, (c) *elicitation*, in which the learner is requested to reformulate their utterance by the teacher, as in 'How do you say that in English?', and (d) *metalinguistic clues*, in which the learner is provided with metalinguistic comments and/or questions by the teacher (e.g. 'We don't say it like that in Japanese'). Unlike recasts, in prompts, instead of supplying alternative reformulations, the learner is offered an opportunity to self-repair their errors.

In research into the benefit of reactive FonF, among these feedback techniques, recasts and prompts have been frequently studied, since they are the most frequently used classroom feedback techniques. In fact, it has been shown that the most frequently used feedback technique by teachers is recasts in various classroom settings, including elementary immersion classrooms in Canada (Lyster & Ranta, 1997) and in Japan (Lyster & Mori, 2006), EFL classrooms at university level (Doughty, 1994) and at senior secondary level (Tsang, 2004), and adult ESL classrooms (Ellis, Basturkmen, & Loewen, 2001; Panova & Lyster, 2002). In these studies prompts were also often observed, but the use of explicit correction

was infrequent.

To provide teacher feedback in the classroom, while attending to other aspects of L2 learning, the teacher requires the ability to make a real-time decision as to whether or not to intervene, what kind of techniques to use to draw the learner's attention, and whether to provide consistent FonF interventions for learner errors. This is not an easy task and the teacher needs sufficient knowledge of effective techniques and skills in using them. I will now examine findings in previous studies to examine the effects of different types of teacher feedback in L2 classrooms.

Lyster and Ranta (1997) compared the effectiveness of different kinds of feedback techniques in four immersion classrooms at the primary level. The effects of each technique were quantified according to the number of learner-generated immediate revisions of their output, which was termed 'uptake' (ibid. 37). It was found that recasts, which were most frequently used by teachers in the classroom, were least effective in eliciting uptake, while the most effective feedback was that given in the form of prompts. For instance, only 18 percent of the recast moves led to uptake, while the uptake rate of the prompts was larger (e.g., elicitation=46%, metalinguistic feedback=45%, clarification request=28%).

Several reasons have been given for the ineffectiveness of recasts in eliciting uptake. One reason is their ambiguity in signalling to the learner that their utterance included errors to be modified (Lyster, 1998b, 1998c; Lyster & Ranta, 1997; Panova & Lyster, 2002). Some learners may perceive recasts simply as alternative ways of expressing the same meaning without any corrective element, while other might see recasts as corrective feedback. Another reason is that '[I]n contrast to self-repair that follows prompts, uptake that involves repetition of a recast does not engage learners in a similarly deep level of processing, nor does it necessitate any reanalysis'(Lyster & Mori, 2006:273).

However, Sheen (2004) has pointed out that the repair rate reported in Lyster and Ranta's study (1997) needs to be interpreted cautiously, since it was based on the total number of recasts rather than the total number of uptake moves following recasts. Not every recast necessarily provided learners with an opportunity to uptake following the recast and 'the rates of uptake following recasts can differ considerably depending on whether learners do or do not have a chance to uptake' (Sheen, 2004:268).

In fact, 'topic continuation occurred frequently following recasts (i.e., 70% of the time) … and the ratio of repair to total uptake was much higher (57%)' (ibid.:268). Added to this, I should point out that in Lyster and Ranta (1997) uptake was used as a measure of acquisition, whereas in fact it was merely a discourse phenomenon observed at that time and the study was not able to show evidence that the observed uptake resulted in L2 development.

There have been several studies which suggest the potential of uptake in the form of a learner's repetition of a recast (e.g. Lyster and Mori, 2006; Mackey, Gass, & McDonough, 2000; Havranek & Cesnik; 2001; Loewen, 2005). For example, a comparative study on teacher-learner interaction in two different instructional settings by Lyster and Mori (2006) has shown that uptake of recasts can be influenced by instructional setting. They examined teacher-learner interaction in two different instructional settings at the primary school level: French immersion in Canada and Japanese immersion in the USA, in both of which content-based L2 teaching was practised. They found a similar distribution of different types of teacher feedback in both settings - recasts were predominantly used (more than half of the feedback), followed by prompts. Explicit correction represented less than 10 percent of all feedback.

On the other hand, the picture was quite different between the two settings in the analysis of the distribution of uptake and repair followed by different types of teacher feedback: more uptake and more accurate response (i.e. repair) was observed in the Japanese immersion (JI) classrooms than in the French immersion (FI) classrooms. Furthermore, '[T]he greatest proportion of uptake and repair in JI settings followed recasts (61% and 68%, respectively), whereas the greatest proportion of uptake and repair in FI settings followed prompts (62% and 53%, respectively)' (291). It was also found that in JI settings recasts resulted in repair as effectively as explicit correction, while this was not the case in the other context, which suggests that recasts served a corrective function in JI settings. Lyster and Mori explained that the differences were brought about because in JI classrooms greater emphasis was placed both on repetition of a teacher's recasts as a means of discourse practice and on accurate oral production than in FI classrooms. It is possible to speculate that following up a recast with repetition can provide opportunities for learners to reflect on the gap between what they have said and the correct sentence that they are repeating after a recast. These results suggest the potential of uptake in the form of a learner's

repetition of a recast; encouraging learners to repeat recasts might serve as a good exercise to facilitate noticing the gap between learner output and reformulation provided in the form of recasts if a classroom's predominant communication orientation is oral accuracy as in the JI setting.

A high rate of noticing of recasts is also reported in an experimental study on learners' perceptions of recasts by Philp (2003). She examined 33 adult ESL learners' accurate immediate recall of recasts received in the preceding oral communication tasks in NS-NNS pairs, where the learners received intensive recasts from the NS on their nontarget-like question forms in five 20-minute treatment sessions given over a period of two weeks. It was found that the learners were able to recall at least 70 percent of recasts, but they did not always notice every detail. It was also found that their recall of recasts was influenced by: the length of the recast (the shorter ones were noticed more); the number of changes made in the recast (ones with a smaller number of changes were more noticed); and their developmental level (recasts including too difficult structures were less noticed). Philp suggests that '[In] part, difficulties in recall may reflect the limitations of working memory'(119) and that the extent to which the learners notice the gap between what they had said and the reformulation given in the recast is modulated by attentional resources.

The importance of considering the influence of context on corrective feedback and learner uptake has also been discussed by Sheen (2004). Using the taxonomy by Lyster and Ranta (1997), the researcher examined distribution of different types of corrective feedback and the rate of uptake following each type of recast in four different instructional settings: French Immersion, Canada ESL, New Zealand ESL and Korean EFL. The results have shown that recasts were the most frequent feedback type across the four settings but were much more frequent in the Korean EFL and New Zealand ESL settings (83 % and 68%, respectively) than in the other two settings. It was also found that in the former two settings the rates for both uptake and repair following recasts were larger than in the latter two.

To sum up, these studies have shown the potential of teacher feedback, both prompts and recasts, in *reactive* FonF, but one needs to remember that the effects of each type of interactional feedback may depend on the instructional settings in which they are provided and other context variables, such as pedagogical focus, learners' proficiency level, age and educational background. Given that the effects

of different types of feedback are influenced by instructional settings, more research is needed before drawing any conclusions about their effectiveness. I should also note that it is not clear from the previous research how much feedback is necessary to facilitate L2 acquisition; the provision of intensive feedback on individuals' errors is not available in normal L2 classrooms and we need more classroom-based studies to examine the effects of interactional feedback.

2.5.2 Proactive focus on form

In proactive FonF approaches, target features are chosen in advance by conducting classroom observation and/or an informed prediction of learner difficulties in using specific L2 features, such as learners' readiness to learn a specific form, which will be further discussed in the next chapter. Various tasks and techniques have been used in previous studies, but they have varied in terms of explicitness in drawing learners' attention to the target form.

The most explicit method is teaching of grammar involving explanation of rules. According to Long's (1991) definition of FonF, instruction containing explicit grammar teaching is seen as focus on formS instruction. In support of using explicit grammar instruction in communicatively oriented classrooms, Lightbown (1998) suggests that brief pre-teaching of grammar is likely to heighten learner expectation that errors in producing targeted forms will be corrected and is effective in raising their awareness of those forms. In fact, in previous studies explicit grammar instruction has been used in combination with one or more FonF tasks and/or techniques at different degrees of explicitness. For example, explicit grammar teaching was combined with input processing instruction (VanPatten and Cadierno, 1993); with consciousness-raising tasks (Fotos, 1994); with enhancement (Lightbown, 1998); and with the dictogloss task (Swain, 1998). In these studies form-focused grammar teaching was separated from communicative activities and yet was integral to the lesson as a whole.

I will now examine the effects of FonF instruction used in these studies.

Input processing instruction

As one way of teaching grammar communicatively, VanPatten (1993, 1996, 2002) has suggested an input-based approach to FonF, which he terms as input processing instruction. In this approach, the learners are first provided with explicit grammar instruction followed by a series of input processing activities intended to encourage understanding of the target structure rather than producing it. Because of the inclusion of the explicit grammar instruction, some researchers have equated this approach with the traditional presentation-based approach, or what Long terms *focus on formS*, (e.g. Sheen, 2002), but VanPatten (2002) argues that it is more appropriate to view input processing instruction as a type of FonF, since it aims both at altering the learners' usual processing strategies to understand input and at encouraging them to establish better form-meaning connections by helping them to pay selective attention to form in the input during input processing.

VanPatten and his colleagues (VanPatten & Cadierno, 1993; VanPatten & Oikkrnon, 1996) conducted classroom studies with adult L2 learners of Spanish and compared learning outcomes of input processing instruction with those of 'traditional instruction' (i.e. production-based practice). For example, in VanPatten and Cadierno (1993) two groups of learners were taught the target form (the positioning of object clitic pronouns in Spanish) and learning gains were measured by two means: a discrete-item written production test (sentence completion) and a listening comprehension test with a pre-post test design.

The "traditional instruction" consisted of explicit grammar teaching and exercises which required learners' oral production of the target structure by 'moving learners from mechanical form-oriented practice (oral and written transformation and substitution drills) to meaningful practice (oral and written question and answer, simple sentence formation), and finally to more open-ended communicative practice (oral and written question and answer, conversation)' (VanPatten & Cadierno, 1993: 230).

Input processing instruction provided the learners with opportunities for comprehension and analysis of input without forced output. The instruction involved the presentation and explanations of the target forms, informing learners of the flexibility of Spanish word order and L1 English learners' natural

tendency to segment L2 Spanish speech input according to their native language's rigid SVO word order and providing examples. The learners then completed several activities (mainly written, and some oral) intended to provide them with opportunities to practise processing the input data to make stronger form-meaning connections.

It was found that the input processing instruction group achieved better results than the other in the two tasks on the posttest, which suggests that input processing instruction which draws learners' attention to form is more effective than the traditional output-based practice in promoting L2 acquisition. Similar results were found in VanPatten & Oikkrnon (1996), however, in both studies a discrete-item measurement was used to assess the effects and no oral production task was employed to compare the effects of the two types of instruction. If the learners had been given an oral production test, there would have been a possibility that the traditional instruction group given many production tasks would have outperformed the input processing instruction group. In fact, many recent studies have failed to show any superiority of input processing instruction over production-based instruction (e.g. DeKeyser & Sokalski, 1996; Salaberry, 1997; Allen, 2000: Kim, 2001; Erlam, 2003).

We should also note that, although VanPatten and his colleagues do not totally exclude production tasks in input processing instruction, VanPatten (1996) claims, like Krashen, that output is not necessary for the development of grammatical competence and may even be detrimental, since it can overload the learner's processing capacity. In this regard, VanPatten's approach is limited, since it lacks opportunities to produce comprehensible output - a necessary condition for L2 restructuring, as we saw earlier.

Consciousness-raising

Consciousness-raising (CR) was an early attempt to integrate formal instruction within a communicative framework (Sharwood Smith, 1981; Rutherford, 1987; Rutherford & Sharwood Smith, 1986, 1988). Proponents of C-R suggested that an instructional focus on a grammatical feature can develop the learner's knowledge of a target form and enhance their awareness of the feature in communicative input afterwards - a process seen as essential for L2 acquisition. A CR task is defined by Ellis (1997:160) as

follows:

> 'a pedagogic activity where the learners are provided with L2 data in some form and required to perform some operation on or with it, the purpose of which is to arrive at an explicit understanding of some linguistic property or properties of the TL'.

Fotos (1994) developed CR tasks in which Japanese learners at university level were provided with grammar problems to be solved in meaningful communication in the target language. The CR tasks required the learners to discover the grammar rule for themselves from the given language data, without being provided with explicit explanations of the grammar structure. In this classroom study, Fotos compared the effects of two types of instruction (*grammar CR instruction* and teacher-fronted grammar instruction) on the learner's use of the target structures (adverb replacement, indirect object replacement, relative clause). Both groups received three 90-minute treatment sessions once a week over three consecutive weeks. The CR task group performed the problem-solving tasks in each treatment session without receiving any instruction on the grammar structure. The grammar lesson group received a formal, teacher-fronted grammar instruction on the target forms, but did not have any discussion of the grammar structure and teacher feedback. Each week, the treatment session was followed by two more 90-minute classes, where the learners received regular instruction consisting of communicative activities which lacked grammatical content.

No significant differences were found between the two conditions in the learners' gains in knowledge of the target features, which was assessed through a written test consisting of grammaticality judgment and sentence production with a pre- and post-test design. The results led Fotos to conclude that the C-R tasks were as effective as the traditional, teacher-fronted grammar lessons for the development of problematic structures. Fotos has suggested three pedagogic advantages in having grammar as the task content. Firstly, grammar CR tasks can increase the validity of communication tasks which teachers tend to consider trivial, especially those in the EFL context in which formal teacher-led grammar teaching tends to feature in many classrooms. Secondly, having grammar as the task content requires learners to use the target structure and to attend to a grammar problem involving the target form seeded in the input in order to solve the task, and therefore, such a task requirement does not allow learners to avoid talking about the target structure. Thirdly, grammar CR tasks can ease the assessment of task

performance by using pre- and posttests on the particular grammar structure.

Thus, Fotos's (1994) results suggest the potential of grammar CR tasks; however, no follow-up test was given to the learners and the study was not able to show evidence that the gains were long-lasting and led to L2 development. Moreover, there are several other issues to be addressed. Firstly, the requirement to discuss the target grammar in the target language in grammar CR tasks may be too demanding for learners with limited L2 oral proficiency. Secondly, because of the inductive nature of grammar CR tasks, learners, especially weaker ones, may need longer to understand the given data including grammar problems written in the target language and to carry out the subsequent problem-solving tasks. Implementation of such tasks in formal education may be difficult due to limited class time. Thirdly, although grammar C-R tasks can be implemented as an interactive communicative task, their grammar-oriented content might lead some learners, especially those who put a high value on correctness, to a grammatically improved but more hesitant, less fluent speech style, and to a reduced exchange of information. Such risks should be taken into account in promoting the learner's well-balanced L2 development.

Input enhancement

In more recent studies, the term 'consciousness raising (CR)' has begun to be called input enhancement, following Sharwood Smith's (1991, 1993) suggestion that CR be renamed because the former problematically implies the possibility of directly manipulating the learner's mental stage, while the latter stresses only the external manipulation of input or task materials in deliberate attempts to enhance better intake of targeted forms in the input. However, although techniques used in both types of approaches to draw learners' attention to grammar were much less intrusive and less explicit than those used in the other types of proactive FonF, techniques used in Foto's (1999) grammar CR tasks were different from those that used input enhancement approaches. For example, while Foto's CR tasks required learners to discuss problems involving the target structure embedded in written texts to achieve the communicative goal, different forms of textual enhancement on drawing learners' attention to specific L2 features used in input enhancement, such as enlarged, bolded, italicised, or underlined input enhancement in written texts, did not require them to do so.

A number of studies have examined the effects of different forms of textual enhancement, and results were mixed. While several studies found positive findings for input enhancement (e.g. Doughty, 1988, 1991), others reported only limited effects of such a technique (e.g. White, 1998), and no significant advantage for input enhancement was reported (e.g. Leow, 1997). White (1998) attributed the limited effect of input enhancement on young school-age ESL learners' acquisition of possessive determiners (his/her) to the lack of explicitness of the enhancement techniques she used (enlarged, bolded, italicised, or underlined input enhancement in written texts) in two learning conditions (a typographically enhanced input flood and the same enhanced input flood in addition to extensive reading and listening to provide further exposure to target rule examples). She emphasised the importance of more explicit approaches to FonF instruction.

I should point out that one weakness of the input enhancement technique is that consciousness is neither observable nor measurable and, so far, no valid way of capturing it has been established. In order to claim that learner's increased proficiency is brought about by 'increased consciousness' through grammar input enhancement approaches, researchers need to provide a valid means of measuring the extent of consciousness engaged.

Task naturalness, task utility, and task essentialness

One serious problem in integrating proactive FonF with communicatively oriented instruction is ensuring that learning difficulty occurs in learner interaction in order to bring it into focus afterwards. Loschky and Bley-Vroman (1993) have suggested three possible ways to make this happen: *task naturalness*, *task utility*, and *task essentialness*.

'In task-naturalness, a grammatical construction may arise naturally during the performance of a particular task, but the task can often be performed perfectly well, even quite easily, without it' (ibid.:132). Indeed, a task involving information exchange about differences between two pictures can naturally lead learners to use various question forms to elicit necessary information from a conversation partner; but it is unlikely that learners will produce questions by using late-acquired Q-forms, such as

question tag (*Q-tag*), which they seldom hear in everyday life, even in L2 classes, especially in the EFL context.

'In the case of task-utility, it is possible to complete a task without the structure, but with the structure, the task becomes easier' (ibid.:132). It is possible to design a task in which the use of the target feature will help learners perform the task more easily and achieve more successful outcomes; but again, learners cannot be expected to use that form voluntarily to complete the tasks if its use is not obligatory.

'The most extreme demand a task can place on a structure is essentialness: the task cannot be successfully performed unless the structure is used' (loc. cit.:132). Indeed, making the use of a target form essential would be most useful for implementing proactive FonF; however, as Loschky and Bley-Vroman (1993) acknowledged, it is very difficult to design such tasks to teach various L2 features without losing learners' focus on meaning. For instance, the use of target form is essential in grammar CR tasks, but learners' attention is largely drawn to form to discover the language structure. One possible way of using task essentialness in proactive FonF is via the dictogloss task developed by Wajnryb (1990), which will be discussed in the following section.

Dictogloss

Dictogloss is a procedure that encourages learners to find out what they do and do not know about the target grammar through analysis and correction in working with peer learners. It consists of four stages: (1) preparation for dictation; (2) dictation; (3) text reconstruction in small groups; and (4) analysis and correction of the reconstructed text. It is intended to provide opportunities for learners to use their productive grammar and they are encouraged to scaffold each other. In dictation 'a short, dense text is read (twice) to learners at normal speed' (Wajnryb, 1990:5) and each student group is encouraged to produce its own reconstructed version, aiming at grammatical accuracy and text cohesion. While in the traditional dictation learners reconstruct the text on their own, in dictogloss learners in small groups are encouraged to scaffold each other by sharing and accepting each other's suggestions in order to create a cohesive text with correct grammar, that approximates the meaning of the original text read by the teacher. In this process, the target language is utilised as a means of communication in a natural way

and the principles underlying dictogloss respect the fundamental precepts of CLT.

Dictogloss has been the subject of a number of studies and commentaries and use of the technique has been widely supported (e.g., Brown, 2000; Jacobs, 2003; Kowal & Swain, 1994, 1997; Storch, 1998; Swain & Lapkin, 1998). As is the case to many other teaching techniques, dictogloss lends itself to 'a host of variations developed by creative second language teachers [...] It is about seeking to create an atmosphere in which students are self -motivated and take an active role in their own learning and that of their classmates and teachers' (Jacobs, 2003:13). For instance, while in the originator's dictogloss (Wajnryb, 1990) learners in small groups are asked to create a cohesive text with correct grammar and other features after dictation, that approximates the meaning of the original text read by the teacher, in Swain's (1998) dictogloss the school children in pairs were asked to replicate the text they had heard as exactly as possible and their attention was drawn more to form[6].

The dictogloss task was used in studies by Swain and her colleagues (Kowal & Swain, 1994; Swain, 1998, 2000; Swain & Lapkin, 1998, 2001) as a *proactive* focus on form technique. Dictogloss has potential relevance to this study and I will now take a close look at how it was used in Swain (1998), which investigated the function of metatalk in learning and the effect of metatalk produced through the dictogloss task on the learners' accuracy in using L2 features targeted in the preceding mini-lesson.

In this study, two classes of 48 French immersion students in grade 8 in Canada performed three dictogloss tasks, one per week for three weeks. Each dictogloss consisted of listening to a passage including target L2 features and the reconstruction of the passage in writing, while sharing notes from listening, and working in pairs collaboratively. The first two focused on number and gender, and the third on verb tenses (*passé composé* and *imparfait*). In each dictogloss, the learners in pairs first heard a recorded text in which forms known to be difficult for them were embedded as targets, and then worked together to reconstruct the text exactly. This was intended to encourage them to negotiate the target forms difficult for them to produce on their own. Prior to each dictogloss, both classes were provided with a short (3-5-minute) review lesson on the target rules 'to heighten students' awareness about an aspect of language that would be useful to them in carrying out the dictogloss' (Swain, 1998:73).

In the first session, both classes received similar instructions about how to carry out the dictogloss after a short review lesson with one difference. Although both classes watched the modelling performed by the teacher and researcher (hereafter, the teachers), showing how to reconstruct the text they had heard from notes taken in listening, in the modelling shown to one class (N=26), the teachers referred to grammar rules in their metatalk while reconstructing the text, while in the other modelling, the students (N=22) watched the teachers discussing L2 forms without referring to metalinguistic terminology. In the second session, similar procedures were repeated. After the student pairs had reconstructed their text, the researcher chose the work of one pair and corrected mistakes, modelling the format appropriate for each group.

In the third session, the procedure used in the second session was repeated without the modelling for data collection. Both classes were given a short (3-5-minute) lesson on the *passé composé* and *imparfait* and then the dictogloss. Each pair was audio-recorded while reconstructing the text which they had heard for the analysis. One week after the third session, the students were given a set of tasks which were of a discrete nature (e.g. dual- or multiple-choice, fill-in-the-blank and translation), except for one (open-ended).

Swain (1998) examined the students' participation in language-related episodes (LREs), defined as 'any part of a dialogue in which students talk about the language they are producing, question their language use, or other- or self-correct' (Swain & Lapkin, 1998:70). It was found that in the third dictogloss the class which was exposed to metatalk with grammatical terminology in the modelling produced more metatalk about the target features (gender and number) during the reconstruction. The student pairs in this class produced an average of 14.8 language-related-episodes (LREs) per pair in each *dictogloss*, while the other class which watched the modelling without metatalk produced only 5.8 LREs. Swain has suggested the potential of collaborative dialogue in promoting output and noticing the gap in learners' interlanguage.

Another finding of this study was that, among a total of 256 problems discussed in pairs in the two groups in the reconstruction task given in the third session, more than half were solved correctly. The

students showed a strong tendency: when they had reached a correct solution for a certain rule through dialogue, they performed accurately in using the same item on the posttest; while when they had reached a wrong solution for a certain rule through dialogue, they tended to be inaccurate in using the same feature on the posttest: '[T]he students tended to 'stick with' the knowledge they had constructed collaboratively' (Swain, 1998:79).

Swain also reported that the students did not always focus on the intended target form. The forms targeted in the third dictogloss were the *passé composé* and *imparfait*, but the students talked more about the forms targeted in the previous two lessons (i.e. number and gender). Swain has suggested that 'students talked about what they needed to talk about according to the state of their own internalised knowledge' (ibid.:77).

Based on these findings, three implications have been suggested by Swain: (a) the need for the teacher to prepare and present tasks in a way that can facilitate metatalk, (b) the importance of teacher feedback on learners' nontarget-like forms produced in their negotiation, and (c) the teacher's understanding of individual differences and of the fact that learners' use of metatalk will depend on the kind of input given in the modelling of how to carry out a task.

The findings in Swain's study clearly show the potential of the dictogloss task as a means of facilitating collaborative dialogue leading to the production of comprehensible output; however, again, this study was not able to show direct evidence that such output could actually result in L2 development.

2.6 Summary

In this chapter, I have reviewed the contribution made by interactionist and cognitive research. The latter, which forms the mainstream of current L2 acquisition research, has brought us clear recognition that L2 acquisition involves important psycholinguistic dimensions and better understanding of acquisitional processes in L2 acquisition by using the input-output computation metaphor. We have also seen the important role of noticing and explicit and implicit knowledge for L2 development. I

have examined various reactive and proactive FonF tasks and techniques used in the previous studies which examined the effects of FonF instruction in communicatively oriented L2 classrooms, which suggest the potential and limitations of both types of FonF approaches.

In the next chapter, I will extend my review to research into the acquisition of English interrogatives, which constitutes an important part of the current mainstream L2 acquisition research, on which the present study draws.

CHAPTER 3
Development in the acquisition of English interrogatives

3.1 Introduction

In the previous chapter, I discussed current approaches to grammar teaching and addressed the need for integrating grammar instruction into CLT. In Chapter 1, I also provided my reasons for choosing interrogatives as targets in my study. In this chapter, in order to position my study in relation to current theory, research, and practice, I will review previous research into learners' L2 development, especially in the acquisition of English question formation.

Firstly, I will discuss a problem in SLA research which investigates development and explain how this problem can be overcome by using established developmental sequences. I then present a 6-stage developmental sequence proposed by Pienemann and Johnston (1987), which has been frequently adopted in previous studies to examine the effects of instruction on L2 development. Secondly, I will summarise findings of Meisel, Clahsen, and Pienemann (1981) and the *Multidimensional Model* of L2 acquisition, on which the 6-stage developmental sequence is based. Thirdly, I will operationalise L2 development and explain my assessment criteria for individuals' developmental stage in this study. Lastly, I will examine findings in previous studies which have informed the research design of my study.

3.2 Stages of development for English question formation

Early work on typical sequences of English interrogative acquisition as an L2 focused on natural, untutored acquisition of questions. For example, Ravem (1974) studied the acquisition of English by two Norwegian children and found preposing without inversion in *wh*-questions before acquiring inversion. Cancino, Rosansky, and Schumann (1978) also studied the acquisition of English by six Spanish speakers' (two children, two adolescents, two adults) over a period of 10 months and found two stages in *wh*-question rule acquisition. At the first stage termed 'undifferentiation', learners did not

distinguish between simple *wh*-questions (e.g., 'Where are you going?') and embedded *wh*-questions (e.g., I don't know where you are going.') and produced both types of questions without inversion. The learners then began to produce simple *wh*-question with inversion sometimes, increasing inversion in *wh*-questions. At the second stage, termed 'differentiation', the learners became able to distinguish between the two types of wh-questions.

Adams (1978) carried out a longitudinal case study over a two-year period and examined the development of *wh*-questions and *yes/no*-questions by 10 Spanish-speaking children. For the development of *yes/no*-question rules, she found three main stages: (a) declarative sentence word order with rising intonation, (b) appearance of a few routine expressions with 'do/can' in *yes/no*-questions (e.g., 'Do you have..?), while kept using canonical word order with rising intonation to signal *yes/no*-questions, and (c) use of *be* inversion in *yes/no*-questions and use of more modals (*will*, *could*, *should*). As to the acquisition of *wh*-question rules, Adams also found three main stages: (a) use of a *wh*-word preceded a declarative sentence (noun phrase plus main verb without a modal) with the subject word either 'I' or 'you', (b) use of *wh*-word followed by a declarative sentence without a modal but frequently with *be* (both as copula and auxiliary), and (c) use of inversion of the subject and *be* (both as copula and auxiliary) in *wh*-questions.

In the next sections of this chapter, I will review findings in research into English interrogative acquisition in more recent studies.

3.2.1 The 6-stage developmental sequence

A number of studies have investigated optimal conditions for promoting L2 development, but many studies simply compared learners' L2 output with the target language and failed to account for whether or not instruction had a positive effect on their interlanguage development – a point also made by Larsen-Freeman and Long (1991:63-69). One way to overcome this problem is to use established developmental sequences in the examination of learners' production before and after instruction. In recent research into L2 development in the acquisition of English question formation, researchers have often adopted the 6-stage developmental sequence and examined the relationship between specific types

of instruction and learners' development, which has provided us with a better understanding of the L2 development of learners considered to be at different stages (e.g. Spada & Lightbown, 1993; Mackey, 1995; Mackey & Philp, 1998; Silver, 1999, 2000; Mackey & Oliver, 2002; McDonough, 2005).

This sequence of development was based on Pienemann and Johnston's (1987) implicational analysis of the acquisition of syntactic and morphological features, the data for which was collected through interviews by Johnston (1985), conducted with 24 adult ESL learners in Australia from different L1 backgrounds. The researchers focused on the different cognitive *processing operations* required to produce English word-order rules and morphemes and categorised them into six developmental stages, based on the extent to which these operations require complex speech processing. The six stages and example sentences for each stage are shown in Table 3.1 on the next page. In the table, word-order rules other than interrogatives are excluded[7].

TABLE 3.1 Developmental stages for English question formation

Stage	Q-form	Examples
1	***Single words*, *sentence fragments*** or ***formulae***	Why? Cat how many?; What color the pendant? Why not?
2	***SVO?:*** Canonical word order with rising intonation.	There are two glasses on the table? You are looking for what shape of plates? There are two women, isn't it?
3	***Fronting: Do-fronting/Fronted-other*** Direct questions with main verbs and some form of fronting.	Does the man have a flower in his hand? How many flat mates are in your flat? How many flat mates you have? Where the cat is? Is there is a ball?
4	***Wh-inversion*** : In *Wh*- & *Q*-word questions the copula and the subject change positions.	Where is the cat? How many people are there in your picture?
	Y/N-inversion: In *Y/N* questions an auxiliary other than 'do' is in sentence-initial position.	Is there a cat in your picture? Have you sent an application form to the university? Can you tell me the name of newspaper?
5	***Aux/Do-2nd*** : An auxiliary verb or modal is placed in second position to a *wh*-/*Q*-word and before the subject followed by a main verb [applies only in main clauses/direct questions].	What can you see through the window? How many flat mates do you have? What is the man doing? Why haven't you sent it yet?
6	***Cancel-inversion*** : *Inversion is not present in subordinate clauses.*	Do you remember when Jill got married? Could you tell me what time I can see you on Friday? Can you tell me if the woman is wearing earrings?
	Q-tag: An auxiliary and a pronoun are attached to the end of the main clause.	In your picture there are two wine glasses, aren't there? The father doesn't live with his family, does he?
	Y/N-Negative : A negated form of Aux/Do is placed before the subject.	Haven't you done the dishes yet?

Notes.

This table is based on Pienemann & Johnston (1987) and Pienemann, Johnston, & Brindley (1988), and an adaptation of their model by Spada & Lightbown (1993) and by Mackey & Philp (1998). The description of each stage is adopted from Spada & Lightbown (1993) and Mackey & Philp (1998). See also Mackey's (1995, 1999) adaptation. The examples are from my study.

We can see from this table that in the first stage, the least complex processing operations are required, while in the sixth stage the most complex ones are needed. At *stage 1*, learners produce single words and sentence fragments and can ask one-word questions, such as 'Sleeping?'. At *stage 2*, learners can produce simple strings of elements based on meaning or information focus, such as *SVO* structures with rising intonation, as in 'You have a cat?'. At *stage 3*, learners are able to identify the beginning and end of a string and to perform operations on an element in these positions, such as shifting it from beginning to end of the string or vice versa, which enables them to produce *Fronting* questions, as in 'Do you play football?'. At *stage 4*, learners can identify an element within a string and move it from the middle of the string to either the beginning or the end. Mastery of these operations allows learners to invert the subject and the auxiliary/copula in *yes/no*-questions (e.g. 'Can you swim?'; 'Is there a cat?'), or in *wh*-questions (e.g. 'Where is my purse?').

At *stage 5*, learners are able to identify elements in a string as belonging to different syntactic categories, shifting them around inside the string, producing *Do/Aux-2*nd questions like 'Where are you playing football?', in which the auxiliary 'are' is placed in the second position of the sentence before the subject in *wh*-questions. At the highest stage, *stage 6*, learners are now able to move an element out of one sub-string and attach it to another element. This stage is characterised by the learner's ability to process across as well as inside strings, which enables them to apply the cognitively most difficult rules, such as *Y/N-Negative* (e.g. 'Haven't you seen a dog?') and those involving subordinate clause operations, such as *Cancel-Inversion* (e.g. 'Could you tell me why you haven't sent an application form to the university?') and *Q-tag* (e.g. 'You can't see a cat in your picture, can you?').

Learners' ability to perform the operations at one stage is a prerequisite for being able to perform those involved in the stage which immediately follows. In other words, production of rules at stage 6 requires learners to have acquired the operations not only at stage 6 but also at all the lower stages. Thus, these operations are implicationally related and their acquisition is constrained by the learner's developing cognitive ability to process them (i.e. the ability to identify and move syntactic elements in grammatical strings). Pienemann & Johnston (1986) have argued that it is not possible for the learner to skip a stage in their development, 'regardless of how much or what kind of instruction he or she receives' (100) and they must go through these predictable six stages in their L2 development, in the

same way as L1 acquisition, irrespective of their age and L1 background (Pienemann and Johnston, 1986, 1987; Pienemann, Johnston, & Brindley, 1988).

It is important to recognise that each operation involves the production of not just one but a number of structures. For instance, the mastery of operations to move an element within a string from the middle of the string to either the beginning or the end, involved in stage 4, allows a learner to produce *Y/N-Inversion* and *Wh-Inversion* questions. However, the fact that a learner has mastered a processing operation does not necessarily mean that she/he can at once apply the operations involved in one stage to all the grammatical structures that could be produced. In other words, there is a possibility that learners can show a gap in their L2 development.

3.2.2 Gaps in L2 development

For example, a learner who had mastered the operations at stage 4 and could produce *Y/N-Inversion* questions predicted for this stage might not be able to produce the other stage 4 rule questions, that is *Wh-Inversion* questions. Such gaps in L2 development were in fact found in a series of studies by Pienemann and Johnston (Johnston, 1987; Pienemann 1984, 1986, 1989, 1992, Pienemann & Johnston, 1986). It was reported in Pienemann, Johnston, and Brindley (1988) that Pienemann (1987) showed that such gaps can be filled through instruction; however, it is not clear from his paper what kind of instruction was given to their subjects. In the previous studies it was also shown that, while working on the processing operations involved in one stage, the learner begins to acquire those associated with the next.

Thus, gaps were found in previous research to examine development in the acquisition of English question formation, but so far no counter-evidence against the sequence of development has been reported and the existence of the sequence is well established in both experimental and classroom-based studies. The major strength of the research using the 6-stage developmental sequence is that it is based on a well-defined theory of L2 acquisition, the Multidimensional Model of L2 acquisition proposed by Meisel, Clahsen and Pienemann (1981). In the following sections, I will discuss the contribution of the model to our better understanding of L2 development.

3.3 The Multidimensional Model of L2 acquisition

This model is based on findings of a project known as ZISA (Zweitspracherwerb Italienischer und Spanischer Arbeiter), which examined, through cross-sectional and longitudinal studies, the naturalistic acquisition process of L2 German word-order rules and lexicon by adult migrants working in Germany and their children (Spanish, Italian, and Portuguese) without formal instruction in German. In the project implicational scaling was used to identify the acquisitional order of various language features in the data collected in informal interviews, which permitted explanation of both underlying systematicity and variability in learners' L2 output.

The basic concept of the Multidimensional Model is that in L2 acquisition there are two types of linguistic structure - *developmental* and *variational* features - which are acquired in different manners. In the 1960s and 1970s the prevailing view in SLA research was that L2 acquisition was a linear and uniform process and the ZISA researchers' view of it as multi-dimensional process was quite innovative at that time. It is now widely accepted among SLA researchers.

3.3.1 Developmental features

Developmental features are inevitably acquired in a particular sequence, from simple structures to complex ones. Examples of such developmental features in German and in English are word-order rules. The ZISA researchers were able to explain why rules were acquired in a particular order by using speech processing operations and demonstrated how the acquisition of each successive feature requires the automatization of the speech processing operations involved in the preceding feature.

3.3.2 Variational features

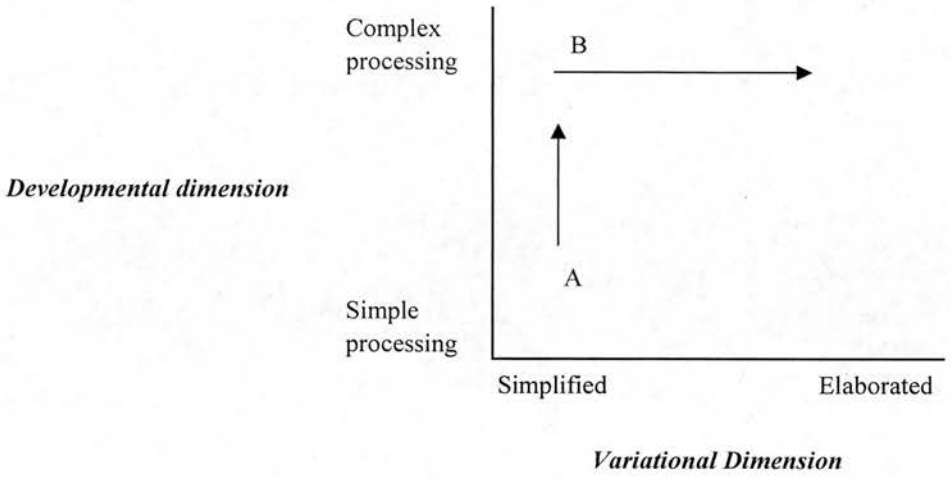
Variational features are not so constrained by the learner's evolving speech processing capacities, although the acquisition of variational features is not entirely random (Meisel, Clahsen, and Pienemann, 1981). The ZISA project identified 12 variational features for German and many of them involved 'deletion' phenomena, such as omission of the verb 'to be', as in 'Ich ø spanisch' (= I Spanish) rather

than 'Ich bin spanisch' (= I am Spanish), and omission of pronoun subjects, which teachers may refer to as 'fossilised' when their occurrences are prevalent and longstanding. They argued that suppliance or non-suppliance of variational features is influenced by various influences, such as socio-psychological learner factors.

In Johnston's L2 English data (1985), deletion of the copula 'be' was also observed among the subjects across various developmental stages and it was suggested that the English copula 'be' is an example of a variational feature (Pienemann & Johnston, 1987). Pienemann and Johnston (1986) argued that all kinds of variational features can be understood in terms of a single kind of tension – tension between *communicative effectiveness* and *correctness* or *standardness*. The researchers also argued that learners who value correctness or standardness in their language will use elaborated language more than those who do not, since they are aware of various problems stemming from the use of simplified language, such as being considered 'uneducated', 'incoherent', or 'unwilling to assimilate', by native speakers in a host country.

Compared with the concept of a developmental hierarchy of stages, the concept of simplification versus elaboration concerning variational features is less reliable because the latter cannot allow us to make predictions about the relationship between the learner's use of variational features and their L2 acquisition. However, the Multidimensional Model offers a framework to explain observed development in using variational linguistic features. For instance, this model can explain several patterns in the learner's L2 development in terms of the two independent dimensions along which the acquisition of an L2 proceeds, as shown in Figure 3.1 on the next page.

FIGURE 3.1 The Multidimensional Model of L2 acquisition
(Adapted from Ellis, 1990:155, based on Johnston, 1987)



A learner in position 'A' can progress along the developmental dimension while remaining oriented towards the instrumental function of variational features; whereas a learner like 'B' can progress along the variational dimension while remaining at a high stage. This suggests that learners can progress at different rates and reach different levels of success in their L2 learning, although they cannot skip developmental stages. Thus, the model has strong predictive and explanatory power for the acquisition of both developmental and variational language features.

3.3.3 Limitations of the model

The Multidimensional Model is a powerful theory of L2 acquisition, but shortcomings of the model and Pienemann and Johnston's extension of the model to English acquisition have been pointed out. Firstly, as Larsen-Freeman and Long (1991) have noted, the ZISA researchers did not explain how learners became able to overcome the processing constraints which normally block learners' progress, and how intake obtained from input is used for restructuring interlanguage systems.

Secondly, the model has been criticised for its use of speech processing constraints to explain the developmental sequence (Larsen-Freeman & Long, 1991; White, 1991; Hudson, 1993). In order to give the model further explanatory power, Pienemann (1998) has provided an account using cognitive

processing limitations and working memory constraints to redefine the model, but the proposed developmental stages remain unchanged.

Thirdly, it has been pointed out that the model is based on research into fairly limited numbers of L2 features and that 'it is not clear how variational features can be identified and, in fact, few examples have been provided, the most frequently mentioned being copula *be*' (Ellis, 1997:60). Although it has been claimed by its proponents that the model is falsifiable, the inclusion of variational features in the model can be problematic, since 'the appearance of a grammatical feature which was predicted not to occur in the model could be explainable in terms of it being a 'newly discovered' variational feature rather than a falsification of the model' (Ellis, R., 1994:400), unless particular features are specified as variational features.

Lastly, as pointed out by Hulstijin (1987), there has been a lack of established quantitative or qualitative criteria by which to judge whether a learner has acquired a specific processing operation. For example, as a means of measurement of 'acquisition', Pienemann and his colleagues counted the occurrences of the suppliance features in obligatory contexts in their original study (Meisel Clahsen and Pienemann, 1981: 112) and proposed the model on that basis. However, as Ellis, R. (1994) pointed out, they later redefined 'acquisition' in terms of 'onset'. The criterion of using the first appearance of a grammatical feature in the data as evidence for acquisition can be problematic because a single occurrence of a feature could be chunk-learned material. For example, an utterance like 'What is he doing?, which is categorised as a stage 5 question, can be produced as chunk-learned material.

The difficulties of distinguishing a formula (i.e. an unanalysed chunk) from an example of successful application of linguistic operations have been discussed in previous studies (e.g. Pienemann & Johnston, 1986, 1987; Pienemann, 1992; Spada & Lightbown, 1993, 1999; Mackey, 1995, 1999; Mackey & Philp, 1998), calling for the use of more rigorous quantitative and qualitative criteria. For example, Pienemann (1992) later used a 'two productive usage' criterion to avoid the problem stemming from the use of formulae in the identification of the *emergence* of a new L2 feature; and Mackey (1995) has suggested a more conservative measure: two productive uses at two different times, which I have adopted in this study. I will return to this in Chapter 5 (Research Design) in order to give further

support for my decision to employ emergence-based criteria, where I will show how I tried to minimise misjudgement in distinguishing formulae from productive use of Q-forms by using a stricter criterion for stage assignment for two posttests than that used for the pretest.

3.3.4 Emergence-based criterion

Thus, the problem of using an emergence criterion has been recognised by researchers, but in SLA research the use of emergence as evidence for L2 acquisition had become another innovation brought in by the ZISA researchers, since by then the use of a mastery criterion was standard in morpheme acquisition studies in both L1 and L2. Unlike a mastery criterion, an emergence criterion does not use accuracy, or production of target-like features, as a means of measuring development. Pienemann (1998) argued that a mastery criterion is not suited for 'a theory which is directed mainly at capturing the systematicity of spontaneous oral production' (148) in the course of interlanguage development. For instance, if one applies Brown's (1973:258) established mastery criterion for L1 morpheme acquisition to research into word order rules, such as Q-forms, there is a high possibility of lack of data to apply the criterion, since his mastery criterion requires 90 percent correct use of an item in obligatory contexts in three successive recordings; in each recording at least five obligatory occasions had to be present. It is likely that the occurrence of word order rules is much smaller than that of morphemes. In fact, Meisel, Clahsen, and Pienemann (1981) reported that sometimes they did not find five obligatory contexts in their subjects' recordings even though the subject was producing a particular L2 feature 100 percent correctly.

Another problem of the mastery criterion is that 'it ignores the insights provided by the interlanguage hypothesis, since only standard-like usage is counted' (Meisel, Clahsen and Pienemann, 1981:112). For example, if we apply Brown's criterion as stated, a learner who fails to supply the correct form of the copula *be* (e.g. supplying 'is' where second person plural 'are' would have been expected) once in five obligatory contexts in one of the three successive recordings is considered not to have acquired this auxiliary at all. (For further examples and discussions, see Meisel, Clahsen and Pienemann, 1981:112).

Thus, the Multidimensional Model and the research based on it have their limitations; however, the

model has both strong predictive and explanatory power for L2 acquisition and no serious questions have been asked about the developmental stages themselves, either in empirical studies (e.g. Pienemann and Mackey, 1993; Mackey, 1995) or in classroom-based studies (e.g. White et. al., 1991; Spada & Lightbown, 1993, 1999). Emergence criteria were therefore adopted in this study.

3.4 Operationalising L2 development

For the purpose of this study, L2 development was operationalised following Pienemann and Johnston (1987), Mackey (1995,1999), and Mackey and Philp (1998). L2 development was considered to have taken place if a learner produced:

(a) *two* productive usages;
(b) of *two* forms at a given developmental stage
(b) at *two* different times.

This is the criterion employed by Mackey (1995, 1999) and Mackey and Philp (1998) for their developmental stage increase assessment and it represents a more conservative measure than the 'two productive usages at one time measure' used in Pienemann & Johnston (1987). This represents the most conservative criterion used in previous studies of English question formation acquisition based on the Multidimensional Model.

I should remind the reader that, although I adopted Mackey and Philp's requirements, the criterion used in my study was more rigorous than their criterion; whereas their studies included three posttests given within four weeks after the last treatment session and their subjects were considered to have increased in developmental stage if they satisfied the requirement on any combination of two of the three posttests, this study included two posttests , the second given 6-7 weeks after the last treatment session. This means that the learners in this study were required to demonstrate a sustained gain over '6-7 weeks', while in Mackey (1995, 1999) and Mackey and Philp (1998) their subjects were required to show a sustained gain of either 'two weeks or four weeks', in other words, a third to a half the time used in this study. My intention was that this more rigorous criterion would allow a stronger claim that stage

development had indeed been sustained. I will further explain the criteria used in this study in Chapter 6 (see Section 6.4, 'Procedures for analysing individuals' gain in developmental stage') and in Chapter 7 (see Section 7.2, 'Effect of the treatment on filling gaps').

3.5 Research into the benefit of instruction

3.5.1 Overview of previous studies

As discussed above, Pienemann and Johnston's (1987) 6-stage developmental sequence has been frequently adopted in SLA research to investigate the benefit of explicit form-focused instruction and/or different interaction opportunities provided in instructions for learners at different developmental stages. Most studies were carried out in an ESL context and, to the best of my knowledge, only one study has been undertaken in an EFL context (McDonough, 2005), as summarised in Table 3.2 on the next page.

The first three studies (White, Spada, Lightbown, & Ranta, 1991, Spada & Lightbown, 1993; 1999) were carried out in classes with about 30 French–speaking children (aged 10-12) in intensive ESL programmes. In these classroom studies, the treatment was provided by the regular classroom teachers. Two of the three studies examined the effects of explicit form-focused instruction; the target Q-forms in White, Spada, Lightbown, and Ranta (1991) were *yes/no* and *wh*-questions with inversion at stage 4 and 5 with inversion and those in Spada and Lightbown (1993) were *yes/no*- and *wh*-questions with inversion at stage 4. In both studies the instruction was followed by activities given by the teachers. In Spada and Lightbown (1993), explicit corrective feedback to the class was also provided while carrying out the activities; in White, Spada, Lightbown, and Ranta (1991) such feedback was not provided. In the other classroom study, Spada and Lightbown (1999), neither explicit grammar teaching nor corrective feedback was given to the students; instead, they were exposed to high frequency comprehensible input of target questions at stage 4 and stage 5.

TABLE 3.2 Studies of the effects of instruction on the acquisition of question formation

	Study context		Number of learners		Learners' L1	Treatment			
			Adults	Children		Explicit instruction	Negative Feedback	Interlocutor in tasks	Targeted Q-form(s)
White, Spada, Lightbown & Ranta (1991)	*Classroom-based*	*ESL*	-	108	French	Provided	(1*)	Peer NNS	Stage 4 and 5
Spada & Lightbown (1993)	〃	〃	-	89	〃	〃	Explicit	〃	Stage 4
Spada & Lightbown (1999)	〃	〃	-	144	〃	Not provided	(1*)	〃	Nontarget-like questions at stage 4 and 5
Mackey (1995, 1999)	*Experimental*	〃	34	-	Mixed	〃	Implicit	Trained adult NS	Nontarget-like questions at any stages
Mackey & Philp (1998:347)	〃	〃	35	-	〃	〃	〃	〃	〃
Silver (1999; 2000)	〃	〃	32	-	〃	〃	〃	〃	〃
Mackey & Oliver (2002)	〃	〃	-	22	〃	〃	〃	〃	〃
McDonough (2005)	〃	*EFL*	60	-	Thai	〃	〃	〃	Non-targetlike questions at stage 5

Notes.

(1*) Negative feedback was not used as an independent variable. The treatment took place in intensive ESL classes, where the instruction was meaning-based and corrective feedback on linguistic forms was rare.

The rest were experimental studies, in which no explicit instruction and corrective feedback was provided (Mackey, 1995; 1999, Mackey & Philp, 1998, Silver 1999; 2000, Mackey & Oliver, 2002, McDonough, 2005). In these studies the learners carried out communicative tasks with a trained NS in one-to-one conversations in three 15-30minute treatment sessions, where some of the learners were provided with intensive implicit negative feedback on their nontarget-like questions at any stage, although in McDonough (2005) the feedback was given to nontarget-like questions only at stage 5.

In the following sections, I will examine findings in the studies from five different aspects:

(a) benefit of modified output in response to negative feedback;
(b) benefit of explicit form-focused grammar teaching;
(c) L1 influence;
(d) optimal timing of form-focused instruction; and
(e) data elicitation methods.

3.5.2 Benefit of negative feedback in promoting modified output

In current research into L2 development, the benefit of negative feedback provided to promote *modified output* has been a primary concern. Researchers have examined which types of feedback and interaction opportunities are effective in promoting modified output. Frequently studied feedback types are: *explicit correction*, *elicitation*, *clarification requests*, *recasts*, *metalinguistic feedback* and *repetition*. Among these six, degrees of explicitness and implicitness vary, with advantages and disadvantages for each type, including the degree of interruption to focus on meaning and flow of communication; that is to say, *obtrusiveness* also differs from one type to another. I will now briefly review the characteristics of each feedback type.

Explicit correction is the most explicit and obtrusive of the six feedback techniques not only because it clearly indicates that what the learner has said is incorrect but because it also provides the correct form. There is little chance for the learner to misunderstand the message, but flow of communication and focus on meaning can be seriously interrupted.

Elicitation refers to techniques used by the teacher to directly elicit the correct form from the learner

without providing it. This gives the learner opportunities to reflect on their output and to notice the gap, but elicitation can also interrupt communication flow and focus on meaning.

Metalinguistic feedback includes 'either comments, information, or questions related to the well-formedness of the student's utterance, without explicitly providing the correct form (Lyster & Ranta, 1997:47). Degree of interruption to focus on meaning and flow of communication is similar to that of *elicitation*.

Clarification requests 'indicate to students either that their utterance has been misunderstood by the teacher or that the utterance is ill-formed in some way and that a repetition or a reformulation is required' (ibid., 47). The correct form is not provided in clarification requests, but the learner is given opportunities for noticing the gap and for reformulating erroneous utterances. This technique is less explicit and obtrusive than the aforementioned feedback types. However, clarification requests are often made by phrases such as 'Huh?' and 'What?' and the learner may not able to identify the problematic features in their utterances.

Recasts refer to the teacher's reformulation of all or part of learners' nontarget-like utterances. Recasts are not introduced by phrases such as 'You mean' and 'You should say' and they are generally implicit and less intrusive than the feedback types discussed above. However, while some recasts are more salient than others in that they may focus only on one word, others may contain grammatical and lexical reformulations and are less salient. Recasts neither identify the learner's erroneous utterances as problematic, nor impose participatory demands on the learner and there is no guarantee that learners will notice what is reformulated by the teacher. Some learners may perceive a recast as a mere repetition of their utterance by the teacher.

Lastly, *repetition* refers to the teacher's repetition of the student's erroneous utterance which is given in isolation. Repetition can be explicit and obtrusive when the error is highlighted by marked intonation or exaggerated pronunciation to draw the student's attention to an erroneous part of utterance. However, like recasts, if their errors involve more than one aspect of L2 use, there is a chance that they may not notice what is wrong with their utterance.

Lyster and Ranta (1997) examined the frequency and distribution of both these six different feedback types used by four teachers in French immersion classrooms at the primary level in Canada and learner uptake (i.e., responses to feedback). They observed 686 teacher turns containing feedback on various aspects of nontarget-like L2 use, such as phonological, and grammatical errors, found that recasts (55%) were most frequently used. The other feedback types were distributed in decreasing frequency as in the following: elicitation (14%), clarification requests (11%), metalinguistic feedback (8%) explicit correction (7%) and repetition (5%). For the students' uptake, two techniques (recasts and explicit correction) involving the provision of correct forms in feedback led to less uptake. For instance, recasts, the most commonly used feedback technique, were least likely to lead to uptake (31%) and explicit correction led to uptake only 50 percent. The other four techniques, which provided opportunities to reformulate an erroneous utterance without providing correct forms, were similar in that they led to more frequent uptake (elicitation=100%, clarification requests=88%, metalinguistic feedback=86%, repetition=78%). The results show that these four techniques were more effective in eliciting student-generated repair (i.e., uptake) than the other two (i recasts and explicit correction).

In this study the researchers further examined types of learner uptake: (a) uptake that results in repair of the error on which the feedback focused (coded as 'repair') and (b) uptake that resulted in an utterance that still needs repair (coded as 'needs-repair'). Among the four techniques which led to more uptake, elicitation and metalinguistic feedback, which provided explicit information about a learner's error, were more powerful ways of promoting student-generated successful 'repair' that involved more than a student's repetition of the teacher's utterance than the other techniques (i.e., clarification requests and repetition of a student's erroneous utterance), which were implicit in providing information about a learner error. Lyster and Ranta suggested the importance of getting learners engaged in reformulating their erroneous utterances through explicit types of teacher feedback.

In research into English interrogative acquisition, the benefit of several types of teacher feedback has also been examined and findings in the previous studies are summarised in Table 3.3 on the next page. Two studies (White, Spada, Lightbown & Ranta, 1991; Spada & Lightbown, 1999) did not directly examine the effects of negative feedback; the treatment in both studies was intended not to provide

corrective feedback on linguistic forms and the teachers' feedback was considered to be implicit. All the experimental studies examined the benefit of negative feedback, such as *recasts* and *clarification requests*, on learners' nontarget-like questions at varying degrees of explicitness or implicitness (Mackey, 1995, 1999; Mackey & Philip, 1998; Silver, 1999, 2000; Mackey & Oliver, 2002, McDonough, 2005).

TABLE 3.3 Findings in previous studies on the effect of negative feedback

	Explicit instruction	Types of feedback	Claims
All studies quoted below	(N.A)	(N.A)	Coprehensible input is necessary but not sufficient. Output through interaction is necessary but not sufficient.
White, Spada, Lightbown & Ranta (1991)	Provided	(1*)	Explicit form-focused instruction is beneficial and necessary.
Spada & Lightbown (1993)	〃	Explicit (error corrections, metalinguistic information)	Explicit form-focused instruction is beneficial and necessary. Explicit corrective feedback may be beneficial and necessary.
Spada & Lightbown (1999)	Not provided	(1*)	Explicit form-focused instruction may be necessary. Metalinguistic information to contrast the difference between L1 and L2 grammar may be beneficial.
Mackey (1995; 1999)	〃	Implicit (*recasts*)	Modified output in response to implicit negative feedback is beneficial and necessary.
Mackey & Philp (1998)	〃	Implicit (*recasts*)	
Mackey & Oliver (2002)	〃	Implicit (*recasts*)	
Silver (1999; 2000)	〃	Implicit (*clarification requests, confirmation checks, comprehension checks*)	
McDonough (2005)	〃	Implicit (*clarification requests*)	

Notes.

(1*) Negative feedback was not used as an independent variable. The treatment took place in intensive ESL classes, where the instruction was meaning-based and corrective feedback on linguistic forms was rare.

In all the studies cited in Table 3.3, it was claimed that comprehensible input is beneficial but not sufficient for helping the learners to advance in developmental stage in the acquisition of question formation. This provides counterevidence to Krashen's Input Hypothesis that comprehensible input is the necessary and sufficient condition for language learning, and hence output and attention to form is not necessary (1980, 1985, 1992, 1994).

In the experimental studies designed to examine the benefit of implicit feedback on use of interrogative forms, either *recasts* or *clarification requests* were used as a means of providing feedback.

Recasts

In recasts, which commonly used by teachers in the classroom, are learners' nontarget-like utterances are reformulated by their interlocutor, as shown below.

NNS: and ah in your picture they children playing?
NS: are the children playing? yes.

(Example from Mackey & Philp, 1998:353)

In this example, the learner did not place an auxiliary before the subject of *yes/no*-question and the NS interlocutor supplied the missing item and reformulated the ill-formed question. Findings from the experimental studies suggest that recasts may be beneficial because they provide positive evidence salient to learners and may serve L2 development (Mackey, 1995; 1999, Mackey & Philp, 1998; Mackey & Oliver, 2002). For example, Mackey & Philp (1998) compared the number of adult ESL learners who increased in developmental stage after the treatment, in which they carried out meaning-focused communicative tasks with a NS interlocutor in five sessions (15-25 minutes/session/ day) on five consecutive days. They found that a greater number of learners who received the treatment with intensive recasts on their nontarget-like questions at any stage (7 of 9=78%) advanced in developmental stage than those who carried out the same tasks without receiving such recasts (1 of 6 =17%). These results suggest the benefit of recasts for L2 development; however, the number of the subjects in both groups was small and studies with a larger sample size are needed to draw a firm conclusion.

The benefit of recasts was also found in Mackey and Oliver's experimental study (2002), which focused on children's acquisition of question formation. Two groups of 22 ESL learners (aged 8-12) from various L1 backgrounds carried out information gap tasks with a trained NS in three 30-minute treatment sessions, but only one group was provided with recasts. It was found that more learners who received recasts (8 of 11=73%) sustained a stage increase than those who did not (3 of 11=27%), suggesting the effectiveness of recasts. The stage increase was observed in the immediate posttest and follow-up test, which contrasts with the results reporting adult learners' development which did not find any immediate effects but mainly delayed effects (Mackey, 1999).

Although these studies suggest the effectiveness of recasts, the limitations of recasts for L2 development have been suggested in empirical studies of learners' perceptions of implicit negative feedback (Mackey, Gass, & McDonough, 2000; Morris & Tarone, 2003). Their findings have shown that learners' nontarget-like morphosyntactic features reformulated in recasts were unlikely to be perceived by learners because of their implicitness. Recasts may be less effective than some other forms of negative feedback, such as clarification requests, since recasts neither identify the learner's erroneous utterances as problematic, nor impose participatory demands on the learner.

Clarification requests

The potential of clarification requests to provide learners with opportunities to modify their output was examined by McDonough (2005) and Silver (1999, 2000). McDonough (2005) carried out the study with 60 adult Thai learners in an EFL context. Open-ended clarification requests (e.g. 'huh?') were provided in two different ways in one-to-one conversations from a trained NS in a laboratory. A group of 15 stage 4 learners received clarification requests made by enhancing the salience of nontarget-like questions through repetition, with stress and rising intonation, as shown in the example (A).

(A) NNS: what angel doing this situation?
NS: *what angel doing? Huh?*
NNS: what is angel doing?

Another group of 15 learners at the same developmental stage received clarification requests without being provided with any information to help them identify the problematic features in their questions, as

in the example (B).

(B) NNS: what happen for the boat?
NS: *what?*
NNS: what's wrong with the boat?

(Example from McDonough, 2005:85-6)

In both conditions, clarification requests were provided only after nontarget-like questions at stage 5. The effectiveness of each condition was compared in terms of the number of learners who moved up in developmental stage. It was found that a greater number of learners (9 of 15 =60%) who received clarification requests and enhanced salience of nontarget-like forms advanced from stage 4 to stage 5 than those in the other condition (5 of 15 =33%). McDonough suggested that *clarification requests* provided with enhanced salience of nontarget-like forms may be more likely to push the learners to notice their problematic features in their output and to produce modified output than clarification requests alone. She has also suggested that clarification requests with the enhanced opportunity to modify may be more effective than other types of implicit negative feedback, such as recasts. I should, however, point out that before drawing conclusions on the effectiveness of the enhanced repetition, one should examine whether or not the proposed enhanced opportunity to modify is more effective than other types of feedback, such as clarification requests following mere repetition of learner output *without* enhanced salience of nontarget-like forms through stress and rising intonation, and mere repetition only.

Silver (1999; 2000) has also suggested the need to provide clarification requests to promote modified output. Silver examined questions produced by 32 adult ESL learners from a variety of L1 backgrounds. They received the treatment without explicit grammar teaching in one of the following three conditions: (a) input condition (N=8) with abundant comprehensible input without producing questions, (b) output condition (N=8) with both abundant comprehensible input and opportunities to produce questions through interaction, but without strong encouragement to produce modified output, and (c) negotiation condition (N=8) with abundant comprehensible input and opportunities to produce questions with strong encouragement to modify output through negotiation of meaning by providing implicit negative feedback on their nontarget-like questions at any developmental stage, which was not given in the other two conditions. The results showed that the negotiation condition was the most beneficial in helping learners increase in developmental stage; in the negotiation condition seven of the

eight learners moved up, while only one learner in the output condition and no one in the input condition moved up. This suggests the importance of the provision of opportunities to produce modified output in response to clarification requests. However, within the three conditions, none of the 24 learners moved up to stage 6 on either of the two posttests, including the nine stage 5 learners. Silver examined the recorded conversations in the treatment and found that few learners produced stage 6 questions and argued that the reason was lack of opportunity to modify nontarget-like stage 6 questions.

This lack of examples of stage 6 Q-forms was also reported in Mackey and Philp (1998). Tasks used in their treatment provided many opportunities to produce questions at various stages, including stage 6, and they had a few stage 5 students ready for stage 6 rules, but when 'compared to the other questions, very few stage 6 questions actually formed the topic of recasts and very few learner utterances could be identified as nontarget-like stage 6 question forms providing the opportunity for a recast' (Mackey & Philp, 1998:352).

To the best of my knowledge, no study has explored learners' developmental stage increase from stage 5 to 6 and no attempt has been made to examine the effects of explicit form-focused instruction targeted on stage 6 Q-forms. One reason for the lack of stage 6 questions produced in the treatment and tests is the difficulty of designing communicative tasks which can elicit stage 6 questions without losing a focus on meaning. And this is also a reason for the lack of research examining the benefit of instruction to focus on stage 6 Q-forms. As a result, little is known about how teachers can help learners to advance from stage 5 to stage 6, and there is a clear need to find answers. The treatment in this study focused on stage 6 rules and I will discuss a possible way to elicit stage 6 questions in tasks later in Section 3.5.6.

To summarise, the results in the experimental studies have suggested that negative feedback, especially clarification requests that facilitate modified output, is effective in promoting development in the acquisition of question formation. I should, however, point out that the learners in these experimental studies received 'intensive' feedback from a NS in three treatment sessions of 15-30 minute one-to-one conversations - difficult to provide in the classroom. There is a need for more research which takes into account the limitations of pedagogical options in the classroom, in order to see how the recommended feedback techniques can be implemented in regular L2 classes.

One possible way to facilitate immediate feedback in the classroom would be to encourage learner pairs to pay attention to their partner's output and to scaffold each other while carrying out a task. SLA research taking into account a sociocultural view of L2 learning has begun to show the potential and limitations of such peer assistance for L2 learning, which will be reviewed in detail in the next chapter. In communicative language teaching, learners are often paired up with a peer to carry out tasks. It is important for teachers to ensure that meaningful peer interaction takes place when instructing the learners to cooperate with each other in pair work. However, in peer conversations learners might feel uncomfortable about giving and receiving some forms of negative feedback, such as correction. There are also limitations on learners' ability to notice their partner's mistakes and to signal them through clarification requests, let alone to provide recasts by reformulating the ill-formed questions. Examination of the potential and limitations of peer assistance is an area which requires more research.

3.5.3 Explicit form-focused grammar teaching

In my experience as a teacher and as a learner, Japanese learners prefer an approach to L2 teaching that draws upon concrete points taught in an explicit, objective way. Added to this, they have few opportunities to use English outside the classroom and tend to expect the teacher to provide both explicit explanation of L2 rules and opportunities to practise what they have been taught through communication activities which lead them from controlled to automatic use of the target structure. In this study, to meet these needs I chose an approach, which provided both explicit form-focused instruction and plenty of opportunities to produce target questions. In the activities used in this approach, learners' attention was drawn proactively to the target form and they were intended to get learners engaged in noticing the gap between the target language and their interlanguage grammar through dialogic conversations for the restructuring of interlanguage. This teaching approach, therefore, represents a compromise between FonF*S* and FonF, with a strong leaning towards the former, in view of the inclusion of explicit grammar teaching at the beginning, and a narrow and intensive focus on one or two pre-selected structures. However, it is not restricted entirely to FonF*S* as defined by Long (1991). I will explain its content further in detail in Chapter 5 (Research Design).

I mentioned earlier that explicit form-focused instruction was found to be effective in facilitating children's L2 development in the classroom (White, Spada, Lightbown & Ranta, 1991, Spada & Lightbown, 1993). For example, in Spada & Lightbown (1993) two experimental classes of 51 learners received approximately 5 hours of explicit form-focused instruction on Q-forms at stage 4 and 5 from their regular classroom teachers. In addition, the next week the learners carried out 4.5 hours of follow-up activities, where they were provided with correction on their nontarget-like questions, again by their regular classroom teachers. The performances of the learners in a communication task (Picture Matching) were compared to those in a control group of 28 children from the same population in a similar intensive ESL programme, in which little form-focused instruction or error correction was provided.

It was found that the overwhelming majority of learners in the two experimental classes either advanced at least one stage or continued to produce more questions at the highest stage they had used on the pretest, indicating the benefits of explicit form-focused instruction. For the control group, the researchers had anticipated smaller effects of the treatment, but it was revealed that more of the students who did not receive the instruction produced questions at higher developmental stages from the beginning than those in the experimental groups. Moreover, many of the control group either maintained or continued to improve their oral production of questions. The analysis of classroom recordings in the control group further revealed that their class teacher focused on form and corrected student errors, although such corrective feedback and form-focused instruction were not in fact supposed to be normal features of the intensive ESL programmes. The researchers concluded that the explicit form-focused instruction and corrective feedback provided within the context of communicative interaction could contribute to L2 development.

Later, these researchers compared the results obtained in the two experimental groups to those found in their 1999 study, in which five intact regular classes of 144 children from a similar population received high frequency exposure to questions at stage 4 and 5 over a 2-week period (1 hour/day on consecutive 8 days) in intensive ESL classes without receiving explicit instruction and corrective feedback on their nontarget-like questions. Tasks used in the treatment in these two studies were similar. It was found that in the 1999 study only seven of the 39 stage 3 learners (18%) and none of the 25 stage 4 learners

went up a stage. Thus, a smaller number of students in this 1999 study advanced in developmental stage than those in the 1993 study, where explicit instruction and corrective feedback were provided[8]. These results suggest the need for the provision of explicit grammar instruction and corrective feedback, even for those learners who are psychologically ready.

3.5.4 L1 influence

Added to these findings, Spada and Lightbown (1999) pointed out negative influence from the learners' L1 in their English question production. In their study, in addition to the oral task, the learners carried out three writing tasks. The analysis of question examples produced in a Scrambled Questions task, which was intended to examine the learners' ability to reconstruct questions from the given words in jumbled order, showed that, despite the frequency of exposure to correct examples of questions at stage 4 and 5 in the treatment, more than half the students constructed ungrammatical questions in answering six of the 20 task questions on both pre- and post- tests. The researchers argued that the production of such mistakes might be caused by word order flexibility in French. In French both *yes/no-* and *wh*-questions can be formed with declarative word order. For example, *yes/no*-questions can be fronted by the phrase 'est-ce que', as in (a), with subject-verb inversion, as in (b), and with a declarative sentence and rising intonation, as in (c).

(a) Est-ce qu'il joue au tennis?
(b) Joue-t-il au tennis?
(c) Il joue au tennis?

English *yes/no*-questions require subject-auxiliary inversion, or *do*-support if no other auxiliary is present. The researchers argued that such word order flexibility in French can allow a francophone to make an assumption that, as in French, there are several acceptable ways to form questions in English. Based on these findings, Spada and Lightbown (1999) have suggested that, in order to help learners notice the difference in the use of specific features between their L1 and L2, there is a need to provide both positive evidence (i.e. what is possible in the target language) and negative evidence (i.e. what is not possible) through explicit comparative metalinguistic explanation.

The present study examines Japanese learners' acquisition of English question formation. Word order in Japanese and in English is very different and it is conceivable that there is negative transfer from their L1 in learning English. For example, in Japanese inversion is not present in both *yes/no*-questions and *wh*-questions and mistakes in realising inversion in English interrogatives are often observed among Japanese learners.

Few studies to examine effects of form-focused instruction on English interrogative acquisition have been conducted with Japanese learners, although two published experimental studies (Mackey, 1999, Mackey & Philp, 1998) included several Japanese learners. In these studies, the Japanese learners' development was consistent with Pienemann and Johnston's claim that the learner must go through the sequence without skipping a stage, irrespective of their L1, but it is apparent that there is a need to test the 6-stage sequence with more Japanese learners before making a firm conclusion.

3.5.5 Optimal timing of form-focused instruction

Pienemann (1989) proposed that instruction is effective only when it focuses on the L2 features at the *next* stage of learners' development, which is now well widely known as the *Teachability Hypothesis*. In order to verify this hypothesis, Pienemann (1984; 1988; 1989) conducted empirical studies with both child and adult learners of German as a second language and examined their changes in developmental stage before and after instruction. The instructional approach used in these studies was highly structured, involving explicit instruction, practice, and production. The findings of the studies suggest that learners benefited most from the explicit instruction when it was targeted at their next developmental stage. He also found that learners who received the instruction targeted two stages beyond their current level not only failed to advance in developmental stage but also had difficulties in producing some variational features which they had previously been able to deal with without problems. However, Pienemann's research has not addressed the key issue of what types of instruction are more effective than others – a point also made by Spada and Lightbown (1999:14).

Two experimental studies (Mackey, 1999; Mackey & Philp, 1998) examined how learners' developmental readiness interacts with their acquisition of English question formation. In Mackey

(1999) 34 ESL learners from various L1 backgrounds were categorised as 'readies' or 'unreadies', based on their developmental stage, and performed meaning-focused communicative tasks with a NS partner in three sessions (15-25 minutes/session/day) on consecutive three days in four different conditions. In the treatment no explicit form-focused instruction on Q-forms was provided. The results showed that the 'readies' who received implicit negative feedback, namely, recasts, through active interaction (n=7) outperformed the 'unreadies' who received the same instruction (n=7). However, the interpretation of these results as supporting evidence for the Teachability Hypothesis is problematic, for two reasons.

Firstly, the learners in these two groups were provided with recasts on any nontarget-like questions at any developmental stage, and the treatment did not necessarily target the learners' *next* stage of development. Secondly, in this study the claim that the 'unreadies' were 'at a lower developmental stage' (Mackey, 1999:568) was made by using the group average developmental stage rather than individuals' developmental stage. For instance, although the group average of these 'readies' was 3.85, this group included learners between stage 3 and stage 5. The group average of these 'unreadies' was 3.14 and the individuals' developmental stage varied from stage 2 and 4. This means that although all 'unreadies' were not ready for stage 6 Q-forms, when it comes to the Q-forms at lower stages, the learners in both groups had different levels of readiness.

Mackey (1999) explained that the tasks used in the treatment were designed to elicit many questions between stage 2 and 6, so that the learners would produce questions at higher level(s) than their current stage, and that there was a good chance of receiving negative feedback on their nontarget-like questions at stage 3 and above. However, the production of stage 6 questions was much less frequent than of those at lower stages, and none of the learners at stage 5 advanced to stage 6 in both the treatment and tests, as it was also the case in McDonough (2005), as mentioned in Section 3.5.2. Again, this suggests the difficulty of eliciting stage 6 questions through tasks.

Mackey & Philp (1998) also examined the potential of negative feedback on L2 development in relation to the learners' readiness to learn higher level Q-forms and found a greater number of 'readies' who received recasts advanced in developmental stage than 'unreadies' (see 3.5.2 for the detailed results). However, the research design was based on Mackey (1995) and the same limitations discussed above

apply in using the results to test the Teachability Hypothesis.

Evidence from classroom research to examine the *Teachability Hypothesis* has produced varying results. In Spada and Lightbown's (1999) study, the learners at stage 2 and 3 received neither explicit instruction nor corrective feedback, but were exposed to the target forms in the input (stage 4 and 5 questions) given by class teachers. The researchers found no greater benefits of instruction for learners at stage 3 who were considered to be 'more ready' to learn stage 4 and 5 rules than those at stage 2. These results conflict with Pienemann's findings which supported this hypothesis. One possible explanation for the different results can be found in the types of input provided in the treatment in these studies. While the learners in Pienemann's studies (1984, 1988, 1989) received highly structured, explicit instruction, those in Spada & Lightbown's study (1999) received no explicit instruction. This suggests that instruction targeting the *next* stage of learners' development can be effective only when explicit form-focused instruction is provided. More research is therefore needed to obtain better understanding of optimal timing and types of instruction to promote interlanguage development.

3.5.6 Data elicitation methods for developmental stage assessment

The developmental stages proposed by Pienemann and Johnston (1987) were based on the analysis of oral interview data, which is considered to provide evidence of learners' ability to use the L2 in real-time communication. In the research into the effects of instruction reviewed above, the developmental stage assessment was also based on oral data[9], but the researchers used communicative tasks rather than interview, so that they could elicit examples of targeted questions within a limited period of time. Information-gap tasks were most frequently used both in the tests and in the treatment to promote interaction while drawing learners' attention to *meaning* rather than to *form*.

However, as we have seen, there has been frequent comment on the absence of stage 6 questions in the treatment and tests, and it is not always easy to design tasks to elicit cognitively more difficult question examples. One possible means of eliciting the target questions can be found in Spada and Lightbown (1993). In this study the learners were given a set of four pictures and instructed to find out by asking questions which one of the four pictures was held by the examiner. The task was presented as a game

and the learners had to guess the correct picture after producing five questions, but if a *wh*-question was not produced spontaneously after two questions, the examiner 'prompted them (e.g. *Can you ask me a question with what?*)' (ibid.:211). From each learner 15 questions were elicited in each test, with a maximum of three *wh*-questions explicitly prompted by the examiner.

The use of such a prompt might draw learners' attention to form while producing questions, and the resulting question examples would be more accurate than those produced without such instructions. Indeed, we cannot deny such possibilities, but I should point out that it is difficult to identify whether or not questions are produced without paying attention to form, since L2 learners can shift their focus from meaning to form, or vice versa, in carrying out a task, as the need arises, irrespective of the task design. Even in fairly meaning-focused tasks, such as free interviews and discussions, learners' attention shifts can occur consciously or unconsciously and it is hard to prove that examples collected in certain tasks are produced without any focus on form. It is important to collect stage 6 question examples for a fuller understanding of the acquisitional process of English Q-forms, which are the target rules in this study, since non-use of stage 6 questions in a given task does not necessarily mean that a learner has not acquired stage 6 Q-forms.

Furthermore, it is also important to take into account two kinds of need in classroom-based studies. On the one hand, as emphasised in Ellis (2003:1), there is 'the need to obtain samples of language use that are representative of how learners perform when they are not attending to accuracy' in real-time communication. On the other hand, it is also necessary to provide learners with opportunities to use in the test the structures they have been taught, even if it is difficult to elicit them from the learners without giving a prompt. In my study, being aware of the merits and demerits of using such an elicitation procedure, I decided to use similar prompts, which will be explained in detail in Chapter 5 (Research Design).

3.6 Summary

In this chapter, I have discussed several theories and their implications for the teaching of English

question formation. The present study draws on the 6-stage developmental sequence which is based on the Multidimensional Model and I have discussed its contribution and limitations, and the methodological issues to be taken into account in the research design of my study. I have operationalised L2 development as 'development was considered to have taken place if a learner produced two productive usages of two forms at a given developmental stage at two different times', which is the most rigorous criterion used in research into the acquisition of English question formation based on the 6-stage sequence. I have also given an overview of the research trends and discussed the findings in previous studies, focusing on the potential of negative feedback and explicit form-focused instruction, L1 influence, optimal timing of form-focused instruction, and data elicitation methods for developmental stage assessment. In the course of this discussion, we have seen the benefit of both reactive and proactive FonF instruction on English interrogative acquisition and the need for studies in the EFL context for fuller understanding of its acquisitional processes.

CHAPTER 4
Sociocultural perspectives on L2 learning

4.1 Introduction

In the previous two chapters, we have seen that our understanding of L2 acquisition owes much to the contribution made by research with a psycholinguistic or cognitive orientation, and indeed, such research has provided us with a better explanation of acquisitional processes which take place in the learner's brain over time. Nevertheless research based on input-output computational models does not take into account what is going on outside the learner's head and provides little explanation as to possible interaction between the individual learner and the social context in interlanguage development. Furthermore, the data used in the experimental studies to discuss the process of L2 acquisition was often collected in laboratory settings and we must be careful not to generalise their findings.

We have also seen a growing interest in understanding how L2 development occurs through interaction in classrooms rather than in experimental settings. Among the classroom studies reviewed earlier, Swain's 1998 study, which examined the effects of collaborative dialogue between students through dictogloss in French immersion classrooms, differs from the others, since it drew not only on *cognitive theory* but also on the *sociocultural theory* of mental activity, rooted in the work of the psycholinguist Vygotsky (1978, 1981, 1987) and others who have adopted a Vygotskyan approach (Lantolf, 1994; Newman & Holzman, 1993; Wertsch, 1985).

Sociocultural theory has been increasingly applied to L2 acquisition research (e.g. Aljaafreh & Lantolf, 1994; Donato, 1994; Frawley & Lantolf, 1985; Gass *et al*, 1998; Kowal & Swain, 1997; Lantolf & Appel, 1994; Lantolf, 2000a, 2000b; Ohta, 2000; Storch, 1999; 2002; Swain, 1995, 1998; Swain & Lapkin, 1998, 2000, 2001; van Lier, 1988, 1996, 2000). It explores the social and cultural influences on a child's development and is a framework for understanding learning and teaching. Drawing on Vygotsky's theories on the co-construction of knowledge, it claims that social interactions and a child's participation in cultural activities are necessary for development.

This study also draws partly on sociocultural theory in my attempt to provide learners with optimal learning opportunities in the classroom and I will now review how sociocultural theory has been applied to learning and teaching in previous L2 acquisition research. I will first introduce the theory and explain its basic concepts. I will then review findings in studies to examine the effects of teacher-learner mediation and peer learner mediation within the sociocultural framework. Then I will examine the role of the learner's L1 in interaction for L2 acquisition.

4.2 Basic concepts in sociocultural theory

Research conducted on L2 learning from a sociocultural perspective is relatively new, as its advocates, Frawley and Lantolf (1985:19), point out:

> 'while Vygotsky's theories on language and thinking have received considerable attention of late in the general psycholinguistic literature, relatively little attention has been given to his theories in second language learning research'.

The most fundamental, distinguishing concept of sociocultural theory is that higher forms of human mental activity (i.e. mental abilities newly developed by children because of their need to communicate cognitive functions) are always *mediated* by socially and culturally constructed mediational means, such as symbols, signs and other psychological tools. It emphasises the role of language as the primary, most powerful cultural means for mediating activities between individuals for learning, for restructuring the mind, and for forming processes of a higher order, self-regulated thought.

In order to deepen my understanding of sociocultural theory and research drawing on it, I will further review the basic concepts which are often referred to in L2 acquisition research. They are: *mediated mind, dialogic interaction, the zone of proximal development* (ZPD), *scaffolding*, and *self and other regulation.*

The immediate environment surrounding children is amorphous. It remains so if they do not receive

help from adults to make sense of the world through the socio-cultural constructs of the society or culture to which they belong. The *mediated mind* refers to a process that allows children to obtain control over higher mental functions, which help them go beyond their limitations in their actions to make sense of the world.

Dialogic interaction refers to interaction between individuals which occurs as they engage in concrete social interaction, which allows the child to connect new knowledge to existing knowledge. In considering the child's mental development, Vygotsky (1981) proposed that any higher mental function necessarily goes through two stages – from an external (social) stage to an internal (mental) stage. At both stages, dialogic interaction serves the child's development. He maintained that the child's earliest development of higher psychological functions appears for the purpose of social interaction through collaboration with other knowledgeable members of the child's culture. Then a process of internalization – the transfer of functions from the social (interpersonal) plane to the cognitive (intrapersonal) domain – takes place, enabling children to control the process of higher mental activities (i.e. thinking).

The *zone of proximal development* (ZPD) is a metaphorical zone which represents a potential for learning. The child's 'development cannot occur if too much assistance is provided or if a task is too easy. Development is impeded both by helping the learner with what she or he is already able to do, and by not withdrawing assistance such that the learner develops the ability to work independently' (Ohta, 2000:52).

Assistance offered in the learning process has an important role for the child's development, which is referred to as *scaffolding*. Lantolf (2000a) reports many examples of scaffolding in the ZPD. For example, the youngest children in an experiment were unable to press a bulb as instructed: to press it when a green light was showing and to stop pressing it when a red light was shown. Older children were able to perform the task when they were talked through the instruction as they carried out the task. This is a good example of assisted performance achieved by *other-regulation*, brought about by an expert guiding the novice through taking responsibility for some parts of the task. Children who needed assistance to achieve the task in the first place become able to perform it with less or no

assistance and develop *self-regulation* over control of the task.

4.3 Sociocultural approaches to L2 acquisition

Research drawing on sociocultural theory focuses on describing the influence of social factors on learners' interlanguage use mainly in the light of general pragmatics and discourse patterns in teacher-learner interaction and in peer learner interaction.

4.3.1 Teacher-learner mediation

Aljaafreh and Lantolf (1994) examined the process of scaffolding, more specifically, the negotiation of corrective feedback, or other-regulation, between the expert (researcher) and the learners in the ZPD in relation to the improvement in L2 composition. Nine adult ESL learners at advanced level received a tutorial in a university reading and writing course, but only three learners' data was analysed. Prior to the tutorial, the researchers determined three learners' grammatical competence and problem areas (articles, tense marking, use of prepositions, modal verbs) by analysing their first composition, so that a researcher could provide collaborative feedback sensitive to the learners' actual level of competence in their ZPD, increasing the degree of the elaborateness of help until the learners showed signs of responsiveness (i.e. self-regulation) towards error.

They created a regulatory scale to determine the learners' microgenetic growth in developmental level through the ZPD, which consisted of a sequence of five transitional levels of development. The five stages were categorised into three general stages of development through the ZPD toward self-regulation and control over the target structures: other-regulation (levels 1 through 3); partial self-regulation (level 4); and complete self-regulation (level 5), as shown below.

Level 1: Learner (L) cannot correct or notice the error despite teacher (T) assistance
Level 2: L notices error alone but cannot correct it without T assistance
Level 3: L corrects error with implicit T assistance
Level 4: L notices and corrects error with no T assistance
Level 5: L no longer makes the error

The researchers observed each student's changes in attending errors in each problem area over the eight-week study and found many examples of microgenetic development. For example, all the students who had been unable to correct errors in producing a certain form without explicit teacher assistance (Level 2) in the first treatment session were able to correct similar errors with implicit assistance from the teacher (Level 3) in the next session. They also reported that different learners often have different ZPDs for the same target language form and argued that 'linguistic forms alone do not provide us with the full picture of a learner's developmental level' (Aljaafreh & Lantolf, 1994:480). For example, a learner who is able to produce a particular structure as a consequence of more implicit forms of feedback is developmentally more advanced than one who needs direct and explicit feedback for the same structure. They suggested that for negotiated feedback to be effective it must be sensitive to the developmental stage within the learner's ZPD where specific features of L2 language are situated.

4.3.2 Peer learner mediation

Ohta (2000) examined interaction between two adult learners of Japanese, Hal and Becky, with different L2 proficiency and different L1 backgrounds, in a university class in the US. They were introduced to a new construction, the desiderative form of the benefactive construction, through teacher-fronted practice making requests using the construction, below.

Watashi wa	sensei ni	empitsu wo	katte	hoshii
I [topic marker]	teacher [dative]	pencil [accusative]	buy	want.

I want the teacher to buy me a pencil.

Then Hal and Becky performed three tasks in pairs: (a) a role play followed by a grammar lecture in English and teacher-fronted practice in forming target sentences, (b) a translation task, and (c) a communicative interview task.

Becky improved dramatically in producing target sentences with Hal's assistance. Ohta argued that Hal's assistance was effective because it was developmentally sensitive, which was 'graduated and contingent'(Lantolf & Aljaafreh, 1995:620). For instance, graduated assistance was observed when

Becky was not able to remember the Japanese word for 'become'. Hal supplied only the first two syllables so that the rest was supplied by Becky with the correct verb ending. Hal's assistance was also contingent; he never interrupted Becky when it was clear that Becky was continuing an utterance, while he did provide help to Becky when she signalled that she was having difficulty in production, even when those signals were subtle. However, he did not supply information every time Becky called for help; he became less responsive to her bids for help, as he seemed to judge that he had already given her enough information about a certain L2 feature in the past, withdrawing support so that she could make her own progress toward complete self-regulation.

Another important finding in Ohta (2000) is that more proficient learners could benefit from interaction with a less proficient learner – in Hal's case, in fluency and the awareness of the state of his own knowledge. Ohta also reported that Hal and Becky were more willing to experiment with newly learnt L2 features in pair work than in a teacher-fronted situation; in collaborative activities they consulted a teacher when linguistic difficulties arose that they were unable to solve collaboratively, when - to put it another way - they found the problem was beyond their ZPD.

These findings led Ohta to conclude that dialogic interaction in the classroom promotes L2 development in the ZPD. As Ohta recognises, her study was small-scale and the findings cannot necessarily be generalised to other contexts, but her detailed examination of learner mediation through discourse analysis was suggestive and has shown the potential for learners to learn from each other through peer interaction. Similar findings have been reported in classroom studies in different learning contexts (e.g. Donato, 1994, with French learners at a university level; Takahashi, 1998, with young children learning Japanese as a foreign language in an mentary school programme in the USA; Garciá & Asención, 2001, with Spanish learners at beginner level at a university; Mennim, 2005, with English learners of Japanese at an intermediate level at a university). These studies suggest the potential of scaffolding in a wide range of classroom contexts.

Another interesting finding in Ohta (2000) was that active scaffolding was observed even in the translation task, which is considered non-communicative and tends to be rejected by language teachers. This suggests that task design does not necessarily determine the productivity of learner interaction – a

point made earlier by Coughlan and Duff (1994) and Swain and Lapkin (1998) - and that learner engagement has primary importance in their development in the ZPD.

Research drawing on sociocultural theory has also shown a link between different interaction patterns and subsequent L2 learning (Storch, 2002; Tin, 2003). For example, Storch (2002) identified four patterns in interaction in 10 dyads of English learners at an intermediate level at a university: (a) collaborative, (b) dominant/dominant, (c) dominant/passive, and (d) expert/novice. Two indices were used to determine the learners' characteristics: (a) *mutuality* (i.e. how they engaged each other's contributions, such as the provision of feedback and willingness to share knowledge) and (b) *equality* (i.e. the number of contributions made by each student and the degree of control of the direction of the task each student took). It was found that collaborative interaction was the most common interaction pattern among the dyads. The analysis of LREs showed that more instances of successful knowledge transfer from LREs to subsequent use were observed among the collaborative dyads, followed by the expert/novice and dominant/passive dyads. The least transfer was found in dominant/dominant dyads. These results suggest the need for the teacher to encourage learners to work collaboratively so that they could benefit from interaction to make progress beyond their potential developmental level in the ZPD.

4.3.3 Mediation through L1

Contrary to the dominant view of an L1 in L2 learning (i.e. language transfer), which posits relatively little positive role for the L1, the sociocultural theory of mind suggests the critical importance of the L1 as a cognitive tool that helps L2 learners scaffold each other. However, in general, the use of L1 in a communication-oriented L2 classroom is considered inappropriate, on the assumption that it will have negative influences on L2 learning.

Swain and Lapkin (2000) examined how a shared L1 (English) was used by student dyads in two French immersion classes. Class D was given a dictogloss task and Class J a jigsaw-storytelling task, after viewing a video-taped mini-lesson on French reflexive verbs and a modelling of task performance in the participating classes. In the two tasks, the same story was used as input; however, the dictogloss task provided a recorded oral text stimulus in order to construct a version of the story in writing, while

the jigsaw task provided a visual stimulus (either odd- or even- numbered pictures from an 8-picture strip cartoon).

For the analysis, the percentage of the use of L1 and L2 was calculated by segmenting the students' utterances into brief exchanges (turns) based on each turn's type of function by using three coding categories. Each includes sub-categories: (a) *moving the task along* (sequencing, understanding, task management); (b) *focusing attention* (vocabulary search, focus on form); and (c) *interpersonal* (off-task, disagreements). It was found that, although the immersion students had attained a high L2 fluency level through the immersion programme, they used much more L1 than the researchers had anticipated. Of the total of L1 use (Class D=21%, Class J=29%), only about 12% of the L1 turns in both classes were *off-task*, while others were used for important cognitive and social functions. For example, the students in both classes made use of the L1 most frequently for *task management* (D=35%, J=43% in jigsaw-storytelling class).

It was also found that the students' use of the L1 for several other functions was influenced by task type. Class D, who performed the dictogloss, produced more L1 turns for *understanding* than Class J, who performed the jigsaw-storytelling task (22% vs. 10%), while producing fewer L1 turns for *vocabulary search* than the Class J (14% vs. 27%). Swain and Lapkin suggested that when Class D worked on the dictogloss task, they needed to work harder than Class J to understand the story content, since no visual aid was provided in Class D. However, this group had the advantage of easier access to vocabulary, provided in the recorded oral text, while the reverse was the case with Class J, who performed the jigsaw-storytelling task.

Moreover, Swain and Lapkin found significant negative correlations between the percentage of L1 turns and the quality of writing of the jigsaw-storytelling task, while this pattern was not observed in the dictogloss task. The study also reported variability in the relationship between the quality of the students' written narratives and their use of L1: some high-achieving pairs made less use of the L1 and others vice versa. The researchers have suggested that a task is but a 'blueprint', and can be interpreted and carried out by learners in a wide variety of ways, depending not only on learners' L2 proficiency but also on factors such as their perception of the task, preference for visual or auditory stimuli, and the

history of dynamics between the students who chose to work together. This echoes the results found in Ohta (2000) and those of Coughlan and Duff (1994) and Foster (1998), referred to earlier.

Based on these findings, Swain and Lapkin (2000) summarised three roles of L1 in L2 learning within the sociocultural theory of mind: (a) to understand and make sense of the requirements and content of the task; (b) to focus attention on language form, vocabulary, and overall organization; and (c) to establish the tone and nature of their collaboration. The researchers concluded that judicious use of the L1 can effectively support L2 learners in carrying out linguistically and/or cognitively complex tasks and 'the use of the L1 should not be prohibited in L2 classrooms, but neither should it be actively encouraged as it may substitute for, rather than support, second language learning' (Swain & Lapkin, 2000: 268).

4.4 Summary

Research drawing on the sociocultural theory has shown the potential of dialogic interaction for co-construction of knowledge not only in teacher-learner mediation but also between peer learners. The classroom abounds in opportunities to work with other learners and we have seen the importance of encouraging learners to cooperate with each other to promote development in the ZPD. Research has informed us of the importance of taking into account social factors in L2 acquisition research, but it often bases its analysis on data collected at a single point in time without focusing on specific L2 features in the interlanguage and tends to fail to accumulate evidence to prove that social factors affect L2 acquisition over time. Tarone (2000) has argued in her review article on cognitive approaches and sociocultural approaches in L2 acquisition research that it is premature for researchers to assert that it is not worth investigating L2 acquisition in its social context. This suggests the need for research pulling both strands together in a single framework by relating cognitive and social factors in the L2 learner.

CHAPTER 5
Research Design

In this chapter I will describe the research design, in sufficient detail to allow other teachers to reproduce the tests and instruction, if necessary. In Section 5.1 I will present the aims of my study. In Section 5.2 I will describe the interrogatives targeted in the instruction given to the students in this study. In Section 5.3 I will provide background information on the students. Section 5.4 comprises an overview of timing of my study. In Section 5.5 I will show two kinds of tests and a preliminary questionnaire used to obtain ideas about the students' general English proficiency, knowledge of the target rules, and background in L2 learning prior to the main study. In Section 5.6 I will present the results of my analysis of the preliminary tests and questionnaire. In Section 5.7 I will describe the content of my instruction to teach the target rules and questionnaires given after each treatment session. In Section 5.8 I will present the tasks used to collect question examples in three tests given before and after the instruction, in order to investigate its effect on individuals' gains. I will also show questionnaires given after each test, intended to explore the students' views on the value of the instruction and on their changes through the treatment. In Section 5.9 I will explain how I piloted earlier versions of teaching and testing materials and what changes I made to improve them. Finally, Section 5.10 contains a summary of the chapter.

5.1 The aims of the study

In previous studies of the effect of instruction on English interrogative acquisition, a main focus has been examination of relationships between specific types of immediate *teacher feedback* on learners' mistakes, such as recasts and clarification requests, and resulting learners' performances. Although this is not the main focus of this study; that does not necessarily mean that I pay less attention to the importance of teacher intervention. However, I am more concerned with exploring, through carefully designed tasks seeded with both the target forms and prompts for learners to assist each other, the potential and limitations of teachers' plans:

(a) to help learners notice the gap between the target language and their interlanguage, and

(b) to promote learners' active involvement in their interlanguage development.

It is a fact of classroom life that it is difficult for the teacher to provide each learner with immediate feedback while performing a task in large classes, let alone intensive feedback as was employed in the previous experimental studies reviewed in Section 3.5.1. In this study, with this classroom reality in mind, I attempted to find effective ways of providing learners with a rich variety of 'affordances' (van Lier, 1996:53) - learning opportunities through conversational interaction in the classroom, for the acquisition of the target L2 structure.

To this end, my pedagogical choice was the use of a *noticing-promotion approach*, intended to promote noticing of the gap between the target language and learners' interlanguage by integrating explicit grammar teaching with noticing tasks designed to provide plenty of opportunities to produce the target rules with a focus on form. This approach was proactive FonF and the students were encouraged not only to pay attention to their own output and their partner's while performing the tasks, but also to help their partner notice mistakes and make self-corrections.

To facilitate noticing and peer assistance, I employed two means in the treatment: (a) videotaped modelling performed by a learner pair and (b) oral encouragement to pay attention to output to notice mistakes and to help the partner make self-corrections. The main purpose of the modelling video was to show the students how to carry out the subsequent conversation task cooperatively in pairs, but I prepared two types of modelling for each treatment session in order to examine (a) whether the different types of input shown in modelling would affect learners' behaviours in the subsequent tasks and (b) effective ways of providing modelling in teaching. The two types of modelling were: (a) a *peer-assistance version* and (b) a *self-correction version*. In both versions, the same learner dyad performed the same task cooperatively from a script and made the same mistakes, which were successfully self-corrected by the same model. The difference between the two versions was that, in the *peer-assistance version*, mistakes were noticed and self-corrected with help from the partner, while the other version featured self-correction without peer assistance. I will show the content of each version in detail in Section 5.7. Few attempts have been made to examine the effect of modelling

performed by learners and I believe this study will prove its potential and limitations.

To summarise, this study has three aims:

to examine the effect of the noticing-promotion approach:

(a) on individuals' gain in developmental stage in interrogative acquisition;

(b) on filling gaps in individuals' acquisition of Q-forms targeted in the treatment in relation to their readiness to learn; and

(c) on learners' perceptions of the usefulness of the treatment; of their changes in producing target questions both inside and outside the classroom.

To investigate the first two aims, I used question examples collected by using two conversation tasks in the three tests - one pretest, one immediate posttest, and one follow-up test. The follow-up test was administered six or seven weeks after the last treatment session in order to examine the sustained effect of the instruction. This is almost twice or three times as long as in Mackey (1995) and in Mackey and Philp (1998), which reported a sustained gain over either two or four weeks in developmental stage in interrogative acquisition.

For the examination of the third aim, I collected the students' self-reports through questionnaires. Involved as both task designer and teacher, I was concerned about the students' own views on the value of practice given in this instruction, which required them to take an active role in assisting each other within a context where immediate teacher feedback was not provided. In addition, I will quote some conversations that provide evidence that the reported changes in the questionnaires had actually occurred during the treatment.

5.2 The target structures

Many Japanese learners seem to have failed to reach a higher developmental stage in the acquisition of Q-forms over six years of secondary education. Moreover, although some learners have managed to reach the highest stage (i.e. stage 6), many seem to fail to acquire all Q-forms, showing gaps in their L2

development[10], so there is a need to explore effective ways of helping them acquire Q-forms.

In the noticing-promotion instruction, I chose the following four higher-level Q-forms as targets - one stage 5 rule and three stage 6 rules:

Stage 5 *Wh-Negative-Aux-2nd* (e.g. 'Why can't we use the washing machine?')
Stage 6 *Cancel-inversion* (e.g. 'Could you tell me if the boy is listening to music?')
Q-tag (e.g. 'The woman isn't writing a letter, is she?')
Y/N-Negative (e.g. Haven't you done the shopping yet?').

I chose these because the participating students, who will be introduced in the next section, seemed to benefit from the instruction focusing on these structures in one or more of the following areas of their L2 learning: (a) consolidation of their knowledge of the target Q-forms, (b) acquisition of the rules, and (c) mastery of elaborated use of these structures for communication. Another reason for my choice was the lack of opportunities to practise these Q-forms in L2 classes. These cognitively demanding Q-forms are usually taught at senior secondary schools in Japan, but most tasks contained in the authorised textbooks are not communicative, and learners tend to fail not only to acquire these rules but also to appreciate the usefulness of being able to ask questions with these rules. Another motivation for my selection was the fact that few studies have explored the effect of instruction on the acquisition of these higher-level Q-forms and little is known about how teachers can help learners acquire them. I hope that findings in this study will result in a better understanding of the need to provide learners with opportunities to practise these rules and of effective ways to teach them.

5.3 Participants

I carried out my main study at two institutions in Japan: Chiba University and Kawatetsu Hospital Nursing College, both in the centre of Chiba Prefecture next to Tokyo.

5.3.1 The University students

At Chiba University, two classes of 30 undergraduates (Class A and B) and a group of 12 volunteers

joined my study in the second semester of the academic year 2001/2002[11] (October 2001-March 2002), as summarised in Table 5.1 below.

TABLE 5.1 Number of students who joined the study

			Female	Male	Total
Treatment Group	University	(Class A)	2	8	10
		(Class B)	10	10	20
	Nursing College		18	0	18
	Total		30	18	48
Comparison Group	University	(Volunteers)	7	5	12

The students in Class A and Class B received the instruction and served as the treatment group (TG), while the group of 12 volunteers simply took the three tests without receiving the treatment and served as the comparison group (CG).

Treatment Group

Class A comprised 10 first-year students (2 female and 8 male) enrolled in the *Material Design for English Teaching Course*. Class B consisted of 20 students (10 female and 10 male) enrolled in the *Literature for English Teaching Course*, 13 of whom were in the third year and seven were in the fourth year, as summarised in Appendix 1. The student names are indicated using pseudonyms. Their ages ranged from between 17 and 22 years. Both courses were held on Mondays; Class A met in the third period and Class B in the fifth. The students in both classes were taking one of these course as part-requirement for an English teaching certificate.

These two courses were originally offered as a one-year course (April/2001-March/2002), but their lecturer was planning to take sabbatical leave in the second semester and the content of both courses had

been covered in the first semester. When I asked this lecturer if she could help me find participants in my study from her students, she suggested the possibility of conducting it with her students on these two courses, since they did not have classes to attend in the class period for each course in the second semester. I took this opportunity and prepared a letter to invite their participation (Appendix 2).

In this letter, I introduced myself and explained that the purpose of my study was to help learners become efficient: (a) in speaking English for communication and (b) in taking an active role in their L2 learning. I also explained the amount of work required for participation in my study, such as regular attendance at a series of four treatment sessions, at three tests given before and after the instruction, and at two kinds of tests which would be administered after signing up. While preparing the letter, I asked an ex-colleague of mine to act as an informal 'research assistant', visiting each class to recruit volunteers and to administer the two tests later in each class on the same day. She visited the two classes at the end of the first semester in mid July 2001 and was introduced to the students by the lecturer. Before the 'research assistant' handed out to each student a copy of my letter, the lecturer encouraged the students to participate in my study. It was however made clear that participation was voluntary, and that their participation or non-participation in my study and their performance in my study would not form any part of their overall assessment of their course work.

Despite the time and effort involved, all 34 students enrolled in the two courses (Class A=12, Class B=22) decided to join my study and took the two tests, which were given later in each class[12], as described later. Two of the 12 students in Class A dropped out after taking the pretest and have not been included as participants in Table 5.1. Of the 22 students in Class B, one was an international student from China and his data was excluded from my analysis, since his first language (Mandarin) might have had a different influence on his acquisition of English Q-forms. I should also note that the data collected from one student in Class B was excluded from the TG in my data analysis, since she wanted to join my study but was busy job hunting at that time and simply took the three tests without attending the instruction sessions. So her data was included in the CG. As a result, the number of students in each class whose data was included in the analysis of the study was reduced to 10 in Class A and to 20 in Class B.

I expected that about one-third of the students might have reached stage 5 in the acquisition of question formation, and that they would most benefit from the instruction because they seemed to be ready to learn all four target Q-forms. The rest of the students seemed to have reached the highest stage (i.e. stage 6), but I predicted that they would also benefit from the treatment since many of them seemed not to have acquired all the target rules.

Comparison group

At the university, I recruited another 12 volunteers (7 female and 5 male) as a comparison group to examine the effect of the treatment (see Appendix 3). I visited another class taught by another lecturer in the same department two weeks before giving the pretest to the TG. I introduced myself and the requirements for participation in my study and asked for their participation. Originally, 13 students signed up, but one student missed the two posttests and her data was therefore excluded from the analysis. These 12 volunteers also belonged to the Faculty of Education and were seeking a certificate in Teaching English at secondary schools as either their first or second major. Six students were in the second year, one student in the third year, and five students in the fourth. The students' ages ranged from between 17 and 22 years.

5.3.2 The nursing college students

In addition to the two classes of university students, one class of 18 first-year students at the nursing college was included in the TG (see Appendix 1). Kawatetsu Nursing College was then a two-year vocational college. Some students entered the college immediately after leaving senior secondary school and others after attending a one-year nursing training course. All had to take two English courses in the first year - one focusing on oral communication (*English 1*) and the other on reading (*English 2*). I used to teach the *English 1* before starting my MSc and PhD studies at the University of Edinburgh. In early 2001, when I was writing up a proposal for my PhD study, the director of studies there asked me to teach the same course in the second semester of the academic year 2001/2002 and I accepted the offer.

From my previous teaching experience there, I predicted that in general their developmental stage in the acquisition of Q-forms would range from between stage 4 and stage 6. When teaching a class with mixed-level students, choosing the appropriate target structures is always a principal concern for the teacher. I eventually decided to use as a main resource the materials I was developing for this study in this class, since I expected that about half the 25 students had reached stage 5 and were ready to learn all target rules. I also predicted that about one-third of them had reached stage 4 and were ready to learn one of the four target rules at stage 5 (*Negative-Aux/Do-2nd*). They did not appear to be ready to learn stage 6 rules, but I thought that they would also benefit from the instruction, since they would be provided with various tasks seeded with plenty of opportunities to produce questions, not only to use the target rules but also easier ones at stage 4 and the other stage 5 rule (*Affirmative-Aux/Do-2nd*). For the remaining students, I predicted that a few had reached stage 6 and that some others had reached stage 3.

There were originally 25 students on the *English 1 Course* that year (all female), but the number of students whose data was included in my analysis was reduced to 18, although all took the instruction and the tests. This is because six students' recordings taken in the pretest were lost, owing to technical reasons, and one student expressed some reluctance about being included in the analysis of my study on her Consent Form (see Appendix 4). This form was given to all students at both institutions after the follow-up test and all students but one gave their consent[13].

5.4 Overview of timing of the study

Before describing the materials developed for my teaching and testing, I will show an overview of the timing of this study, which is summarised in Appendix 5.

In Edinburgh

I started the materials development for the tests and instruction in May 2001 and completed revision by the end of October. The materials I prepared were:

(a) two kinds of preliminary tests and a questionnaire,
(b) teaching materials for the instruction in the main study,
(c) testing materials for the three tests given in the main study, and
(d) questionnaires given after each treatment session and each test in the main study.

I prepared most of the materials and piloted them in Edinburgh before leaving for Japan in the middle of October 2001 to carry out the main study.

I recruited three female Japanese as examiners for the three tests in the main study in April, and then prepared the letter to invite the university students[14]. In May 2001 I started preparing two kinds of tests, a *Dictation Test* and the *Grammaticality Judgment Test*, to assess the students' general English proficiency and knowledge about the target Q-forms, which will be discussed later. In addition, I prepared a questionnaire to examine individuals' background in English learning. I sent these materials to the aforementioned ex-colleague in Japan at the beginning of July, so that she could visit each institution as 'research assistant' to recruit participants from the two university courses and administer the preliminary tests at the end of the first semester, in the middle of July of that year. The test papers and questionnaires were returned to Edinburgh for analysis at the beginning of August.

While preparing the materials for the preliminary tests, I also started developing materials for teaching and testing in the main study. In order to pilot their earlier versions, I recruited Japanese volunteers studying English in Edinburgh in the middle of August. The piloting was held on three different dates – twice in August and once in early September. While doing so, I prepared two versions of the modelling video for the four treatment sessions in the latter half of August.

By the time I finished analysing the two preliminary tests late in August, I had developed most of the teaching and testing materials for the main study. It would have been ideal to delay the materials development for the main study until I obtained the results of the preliminary tests, but I had to do these two tasks simultaneously with the main study in Japan near at hand. I then revised earlier versions of the materials for the main study and left for Japan in the middle of October.

In Japan

I had about three weeks before giving the first pretest and planned to do the following three things:

(a) recruit volunteers at the University, who served as the CG;

(b) give the examiners two training sessions to familiarise them with both the materials used in the pretest and its procedures; and

(c) book rooms for the three tests at each institution and arrange the necessary equipment to record the students' test performance.

In addition, before giving the pretest at each institution, I met the students in each class. At the nursing college, I taught one 90-minute class, in which I introduced myself, the content of the instruction, and the schedule for the treatment sessions and testing. I then gave the class a speaking task without any focus on Q-forms. I first asked the students to find a partner, with whom they were to carry out the tasks throughout the four treatment sessions. I then instructed the students in pairs to introduce each other in English. I wrote down a few items for self-introduction on the whiteboard in front of the class, such as name, favourite food and drink, films, and what they had done last weekend. I then provided modelling. The students were told to take notes while listening to their partner's self-introduction, so that they could introduce their partner to the class later.

At the university, I was asked by the lecturer on her behalf to administer a written examination in each class, which she had prepared for the assessment of her students' course work. After giving the test in each class, I had about 20 minutes to spend with the students. I asked them to find a partner with whom they were to carry out the tasks in the treatment sessions and to write their names in pairs on a sheet of paper.

The main study consisted of a series of four treatment sessions, given once a week over four consecutive weeks in November and December 2001, plus the three tests. At each institution, the pretest was administered one week before the first treatment session in November, and the posttest one week after the last treatment session in December. After the posttest, the students at both institutions had a winter recess for about two weeks. The follow-up test was in the middle of February 2002, six weeks after

the last treatment session at the nursing college, and seven weeks later at the university. Between the posttest and the follow-up test, the students at both institutions received no further instruction focusing on the target rule, although I gave one lesson to each class at both institutions[15]. For the 12 volunteers in the TG, I made an appointment with each student and gave each test within one week after testing the students in the TG. After completing the main study, I returned to Edinburgh to analyse the data[16]. In the next section, I will show the two preliminary tests and analysis results.

5.5 Materials development for the preliminary tests

As mentioned earlier, I prepared two tests and gave them to the students in the TG before the main study. They were the *Dictation Test* and the *Grammaticality Judgment Test*. The *Dictation Test* was intended to assess the students' general English proficiency and the *Grammaticality Judgment Test* to examine their knowledge of the targeted Q-forms. I also prepared a questionnaire to see in advance the students' general background in English learning. At each institute, one 90-minute period was available to administer the tests and questionnaire in each class and I designed them to be completed within that time.

5.5.1 Dictation Test

The Dictation Test was taken from Fountain & Nation (2000). They introduced four equivalent forms of vocabulary-based graded dictation (A, B, C and D) as a placement test. There were two reasons for my choice. First, although dictation tests measure listening, but not speaking directly, dictation involves 'active and creative processing' (Oller, 1979:41) to produce meaningful sequences of language in relation to extralinguistic contexts' (Oller, 1979:263), and it has the advantage of correlating highly with other tests (Weir, 1990). In fact, Fountain and Nation's study (2000) demonstrates high reliability and validity of vocabulary knowledge, which would affect a wide range of language skills including grammar. For instance, learners are required to recognise typical collocates of the word and the grammatical patterns and functions it displays in a sentence. Second, it can be carried out in a short time (about 10 minutes), as explained below, and its marking is systematic, with fewer problems than in marking normal dictation. While marking Fountain and Nation's dictation tests requires markers to check 90 key words in the text and they can ignore mistakes with the regular *–s, -es* and *-ed* suffixes, marking a normal dictation involves examination of various aspects of language use, such as vocabulary and grammar including concord, and markers need to keep applying its detailed evaluation criteria. In other words, markers' failure to do so can result in different scores.

I chose Test C for my study and recorded its text on mini disk. It was read by a native English speaker at natural speed (see Appendix 6 for the script). The text consists of five sections: the introductory

paragraph (15 words) and four main paragraphs (37-47 words for each). The introductory paragraph included 10 key words and in each main paragraph 20 key words were underlined in the script. The length of the dictated chunks, marked by a slash, increases from one section to another. Grammar complexity and vocabulary difficulty also increases as the test proceeds from one paragraph to the next. The text was recorded with pauses in and after each sentence, to give the students time to write it down on the answer sheet provided (Appendix 7). The text was read only once, but the students were given two minutes after hearing the whole text, so that they could make changes to their answers. In order to help the students understand how to take this test, I prepared an instruction sheet in Japanese and gave an explanation of the dictation test and the procedures (see Appendix 8).

The marking of the test is made simple because only the underlined key words in each paragraph are assigned points. One mark is given for each key word, making a possible maximum of 90 marks for the whole test[17]. Following Fountain and Nation (2000), I ignored mistakes with the regular *–s, -es* and *-ed* suffixes and awarded a mark if the rest of the word was correct.

5.5.2 Grammaticality Judgment Test

The Grammaticality Judgment Test was a written test intended to obtain baseline information about individuals' current knowledge of the target rules in two respects: (a) appropriate word order and (b) use of pragmatically appropriate questions. The test consisted of 50 pairs of sentences (Q1-Q50), of which 40 pairs involved question formation, while the remainder were distractors. They were printed on a set of three double-sided sheets; with the instructions and two sample questions on the front page and the 50 pairs of sentences (see Appendix 9) on the other pages. For each pair of sentences, the students were asked to choose one of the five responses indicating whether both sentences (A and B) were appropriate, neither sentence was appropriate, only sentence A was appropriate, only sentence B was appropriate, or refrain from judging by choosing 'I don't know'. They were not asked to correct sentences which they judged incorrect. For instance, to get the right answer for Q5 involving *Cancel-inversion* reproduced below, the students had to realise that the rule was appropriately applied only in sentence A and needed to tick 'only sentence A was appropriate'. One point was awarded if a student got the right answer. It should be noted that American spelling was used in testing and teaching materials used in this study,

since it was used in the authorised textbooks for schools.

Q5
A. Can you tell me how many spoons you need?
B Can you tell me how many spoons do you need?

Thus, the two sentences in each pair were intended to examine the students' knowledge about one Q-form, but 10 of the 40 pairs involved two different Q-forms. Both sentences in these 10 pairs were appropriate in terms of *word order*, but one of them in each pair was pragmatically inappropriate to convey the intended message specified on the task sheet. For example, in Q26 the students had to invite an imaginary friend to stay in their house, as cited below.

Q26. You are inviting your friend to stay with you in your house this weekend.

A. Don't you come and spend the weekend with me?
B. Why don't you come and spend the weekend with me?

Sentence A involved *Y/N-Negative* and sentence B involved *Negative-Aux/Do-2nd*, in which each rule was applied appropriately in terms of word order, but the students had to realise that the right answer here was 'only sentence B is appropriate', because the use of *Y/N-Negative* questions with 'Don't' in the given context was pragmatically inappropriate. If a student ticked the correct answer (i.e. 'only sentence B was appropriate'), he/she was given one point for each sentence and the value '1' was assigned to the two Q-forms. If a student ticked 'Both A and B were appropriate', no point was given although for sentence B the student's answer was correct. This is because there was a possibility that the students able to tell that both sentences were formed grammatically in terms of word order could take a good guess that one of the two options should have been correct and choose this answer to play it safe, without possessing the knowledge necessary to judge whether the use of Q-form in each sentence was pragmatically appropriate. I summarise the Q-forms examined in each pair of sentences in Table 5.2 on the next page. It should be noted that, in this table, for the 30 pairs in which the same Q-form was seeded in both sentences, the column for sentence (b) is left blank.

TABLE 5.2 Q-form seeded in each pair of sentences

No.	a.	b.	No.	a.	b.
1	*Fronting*		26	*Y/N-Negative*	*Nebative-Aux/Do--2nd*
2	*Wh-Inversion*		27	*(Distractor)*	
3	(Distractor)	(Distractor)	28	*Cancel-Inversion*	
4	*Affirmative-Aux/Do-2nd*		29	*Y/N-Negative*	*Y/N-Inversion*
5	*Cancel-Inversion*		30	*Fronting*	*Y/N-Negative*
6	*Affirmative-Aux/Do-2nd*		31	*Y/N-Negative*	*Q-tag*
7	*Affirmative-Aux/Do-2nd*		32	*Q-tag*	
8	(Distractor)	(Distractor)	33	*Y/N-Negative*	
9	*Q-tag*		34	*(Distractor)*	(Distractor)
10	*Affirmative-Aux/Do-2nd*		35	*Affirmative-Aux/Do-2nd*	
11	*Wh-Inversion*		36	*Q-tag*	
12	(Distractor)		37	*Q-tag*	
13	*Fronting*	*Y/N-Negative*	38	*(Distractor)*	(Distractor)
14	*Affirmative-Aux/Do-2nd*		39	*Y/N-Inversion*	
15	*Y/N-Inversion*		40	*Q-tag*	
16	*Affirmative-Aux/Do-2nd*		41	*Cancel-Inversion*	
17	*Y/N-Inversion*		42	*Fronting*	*Q-tag*
18	(Distractor)	(Distractor)	43	*Q-tag*	
19	*Y/N-Inversion*		44	*(Distractor)*	(Distractor)
20	*Q-tag*		45	*Cancel-Inversion*	
21	*Y/N-Negative*	*Fronting*	46	*Y/N-Negative*	*Y/N-Inversion*
22	*(Distractor)*	(Distractor)	47	*Y/N-Inversion*	
23	*Affirmative-Aux/Do-2nd*		48	*Fronting*	
24	*Y/N-Negative*	*Y/N-Inversion*	49	*Fronting*	
25	(Distractor)	(Distractor)	50	*Y/N-Negative*	*Y/N-Inversion*

This test included question examples between stages 3 and 6. The total number of each Q-form included in the 40 pairs was: seven *Fronting* questions at stage 3; nine *Y/N-inversion* questions and two *Wh-inversion* questions, both at stage 4; one *Negative-Aux/Do-2nd* question and eight *Affirmative-Aux/Do-2*nd questions; both at stage 5, four *Cancel-inversion* questions; nine *Q-tag* questions and 10 *Y/N-negative* questions, three of which were at stage 6. The total number of each Q-form tested was not the same, since the main purpose of the test was not to examine individuals' acquisition of each Q-form, but to obtain baseline information about the students' current knowledge of the rules for the development of teaching and testing materials and a rough idea of individuals' developmental stage in the acquisition of question formation.

In order to encourage the students to answer all 50 questions and to ensure that the two preliminary tests and the questionnaire given after the tests would be completed within the given time (90 minutes), I made a recording of the instructions and the 50 pairs of sentences on audiotape and used it with the written texts. All the instructions were recorded in Japanese and the students were encouraged to

answer the questions one by one, while listening to each pair of sentences recorded by a native English speaker at natural speed. A silence of five seconds was recorded between two pairs of questions, allowing them time to think and choose one out of the four responses. After hearing the recorded instructions, the students heard two pairs of sample questions one by one and were instructed to choose one of the five responses. After answering each sample question, they heard the right answer and a short explanation recorded in Japanese. They were not allowed to turn over the front page until told to do so.

5.5.3 Procedures for the two preliminary tests

The two preliminary tests were given one after another without a break in each class at each institution at the end of the first semester in July 2001 - about three months before the main study[18]. The Dictation Test took about 20 minutes and the *Grammaticality Judgment Test* about 30 minutes. At the university, before the tests, my research assistant gave the students a copy of my letter (Appendix 2), in which I invited their participation[19]. At the nursing college, before administering the two tests, the assistant was introduced to the class by the director of studies as a representative of their future teacher (i.e. myself), currently working in Scotland. Then the assistant gave the students a copy of my letter (Appendix 10), in which I introduced myself and explained that the purpose of the preliminary tests was to examine their general English proficiency for teaching and testing material development for the course. The letter included no invitation for participation in my study because they would be taking my classes as a routine part of their course work. For the group of 12 volunteers at the university, each student was given the Dictation Test by myself immediately after the pretest, but the Grammaticality Judgment Test was not given because it was intended to collect data for the teaching materials development.

In the next section, I will present an analysis of the two preliminary tests and explain how I used their results in the teaching and testing materials development.

5.6 Analysis of the preliminary tests

5.6.1 Results of the Dictation Test

The Dictation Test consisted of one introductory paragraph and four body paragraphs. I calculated average marks for each paragraph for each class and I present the results in Table 5.3 below.

TABLE 5.3 Average scores on the *Dictation Test*

		Intro.	Paragraph 1	Paragraph 2	Paragraph 3	Paragraph 4	Total
		(10)	(20)	(20)	(20)	(20)	Full mark (90)
Treatment Group							
University (N=30)	Class A (N=10)	9.6	10.7	10.5	7.2	6.8	44.8
	Class B (N=20)	9.3	14.0	11.9	9.7	8.2	53.0
	Total	9.4	12.9	11.4	8.9	7.7	50.3
Nursing College (N=18)		4.1	1.2	0.9	0.5	0.3	6.9
Comparison Group							
University (N=12)		9.4	14.8	12.9	10.3	8.4	55.8

For the two classes of university students in the TG, the total average mark was higher in Class B (53.0) which consisted of the third- or fourth-year students, than in Class A (44.8) which comprised exclusively the first-year students. This suggests a possibility that the difference in their general English proficiency between the two classes in the TG could affect their oral performance in the three tests, given that dictation tests correlate highly not only with listening test scores but also with other language skills. The total average of these university classes was 50.3, lower than that observed in the comparison group (CG= 55.8), suggesting that those in the CG had an advantage over the those in the TG in performing the tasks in the three tests used in the main study. The lowest and highest average marks among the students in the TG were 28 and 71, and those in the CG were 46 and 78.

The nursing college students got a very low total average mark (6.9 out of 90), which suggests that their

English proficiency was much lower than I had expected and that many of them would have difficulty in understanding instructions in English. The lowest and highest average marks among these students were three and 20. Few of them had taken a dictation test before[20] and their unfamiliarity with this type of test might be one reason for their low score.

5.6.2 Results of the Grammaticality Judgment Test

The Grammaticality Judgment Test was given only to the students who received the instruction. I calculated each student's success rate in choosing the right answer for each Q-form in the 40 pairs of sentences by dividing the number of correctly answered task questions involving each rule by the total number of task questions in which each rule was seeded. For example, there were nine task questions in which *Q-tag* was applied in one or both of sentences (Q9, Q20, Q31, Q32, Q36, Q37, Q40, Q42 Q43). If a student got the right answer only in two task questions, that student's success rate for this rule was 22 percent. Then I further calculated the average success rate for each Q-form for each class and summarised its results in Table 5.4. In the table below, the four target rules are marked with an asterisk.

TABLE 5.4 Percentage of students who answered correctly in the *Grammaticality Judgment Test*

Stage	Q-form	University		Nursing college
		A (N=10)	B (N=18)	(N=17)
3	*Fronting*	69	70	49
4	*Y/N-Inversion*	68	75	44
	Wh-Inversion	100	94	38
5	*Affirmative-Aux/Do-2nd*	90	85	32
	* *Negative-Aux/Do-2nd*	70	61	6
6	* *Cancel-Inversion*	73	65	18
	* *Q-tag*	70	74	38
	* *Y/N-Negative*	55	52	31

For the two university classes in TG, no notable differences were found for each rule. For instance, for the four target rules, the success rate for each in Class A and Class B was: *Negative-Aux/Do-2nd* (70%:61%), *Cancel-Inversion*, (73%: 65%), *Q-tag* (71%: 74%), and *Y/N-Negative* (55%:52%). For

the non-target rules at lower developmental levels, a similar trend was observed: *Fronting* (69%:70%), *Y/N-Inversion* (68%:75%), *Wh-Inversion* (100%:94%), and *Affirmative-Aux/Do-2nd* (90%:85%). These results suggest that the students in the two classes had similar amounts of knowledge about these rules.

Thus, while the average success rate for each rule observed in the two classes of university students ranged between 52 percent and 100 percent, that in the nursing college was much lower, ranging between 6 percent and 49 percent. For example, as for the target rules, their average success rate was: *Negative-Aux/Do-2nd* (6%), *Cancel-Inversion* (18%), *Q-tag* (38%), and *Y/N-Negative* (31%). As to the non-target rules, it was: *Fronting* (49%), *Y/N-Inversion* (44%), *Wh-Inversion* (38%), and *Affirmative-Aux/Do-2nd* (32%). It should be noted that the data from one nursing college student (Ai) was excluded from the analysis, since she ticked 'I don't know' for all 50 questions[21].

I sorted the average success rate for the eight rules at each institution in descending order, as shown in Table 5.5 below. The most well-answered Q-form at each institution is listed at the top. The number before each Q-form indicates its developmental stage.

TABLE 5.5 Percentage of students' success in choosing the right answer

University (N=28)		(%)	Nursing college (N=17)		(%)
4	*Wh-Inversion*	97	3	*Fronting*	49
5	*Affirmative-Aux/Do-2nd*	88	4	*Y/N-Inversion*	44
* 6	*Q-Tag*	73	4	*Wh-Inversion*	38
4	*Y/N-Inversion*	72	* 6	*Q-Tag*	38
3	*Fronting*	69	5	*Affirmative-Aux/Do-2nd*	32
* 6	*Cancel-Inversion*	69	* 6	*Y/N-Negative*	31
* 5	*Negative-Aux/Do-2nd*	66	* 6	*Cancel-Inversion*	18
* 6	*Y/N-Negative*	53	* 5	*Negative-Aux/Do-2nd*	6

For the university students in TG, the best-answered Q-forms were stage 4 *Wh-inversion* (97%) and stage 5 *Affirmative-Aux/Do-2nd* (88%). The next best-answered rules were stage 6 *Q-tag* (73%) and stage 4 *Y/N-Inversion* (72%), followed by stage 3 *Fronting* and stage 6 *Cancel-Inversion* (both 69%).

The other two target rules were ranked towards the bottom of the table: *Negative-Aux/Do-2nd* (66%) and *Y/N-Negative* (53%). This suggests that more students made mistakes in judging the correct use of these target rules than that of the other rules. I predicted from these results that many students' use of higher-level Q-forms in spontaneous conversation would not be as good as that in the written test.

Furthermore, in marking I found that many students made mistakes in judging the pragmatically appropriate use of the two different negative question sentences in Q26, shown below. As mentioned earlier, only the second sentence is appropriate in the given context (i.e. to invite someone to one's house) and they had to realise that the use of *Y/N-Negative* with 'Don't' in the first sentence was pragmatically inappropriate. However, 10 out of the 28 university students failed to choose the right answer.

A. 'Don't you come and spend this weekend with me?'
B. 'Why don't you come and spend this weekend with me?'

I also observed that many students failed to notice agreement mistakes in tag questions. For *Cancel-Inversion*, many students made mistakes in judging the correctness of indirect questions in which inversion was present in the subordinate clause. These observations suggest a need to focus on these aspects of usage in the instruction.

On the other hand, at the nursing college, three rules at stage 3 and stage 4 were relatively well-answered (stage 3 *Fronting*=49%, stage 4 *Y/N-inversion*=44%, stage 4 *Wh-inversion* and stage 6 *Q-tag*=both 38%), but the other rules at stage 5 and stage 6 were ranked toward the bottom of the table. The average success rate for these structures was: stage 5 *Affirmative-Aux/Do-2nd* (32%), stage 6 *Y/N-Negative* (31%), stage 6 *Cancel-inversion* (18%), and stage 5 *Negattive-Aux/Do-2nd* (6%). I had expected a better success rate for the two stage 4 Q-forms and one of the two stage 5 rules (*Affirmative-Aux/Do-2nd*) and learnt from these results that the class of students I was going to teach in the main study might have less knowledge about these stage 4 and stage 5 rules and lower English proficiency than those whom I had taught there in the past.

5.6.3 Changes made to the teaching and testing materials

By the time I finished marking the two preliminary tests, I had started piloting early versions of teaching and testing materials with 11 volunteers studying in Edinburgh. The analysis of the two preliminary tests showed that many nursing college students' general English proficiency and knowledge about the target rules was not as good as I expected, but there was no time to make major changes to my research design, such as the target rules. I decided to proceed with my study without making major changes, but in order to help the weaker students in the treatment and in the three tests, I made some changes to the teaching and testing materials. For instance, I decided to give the four treatment sessions mainly in Japanese; that does not necessary mean that I paid little attention to the importance of giving instructions in the target language; rather it was a necessary measure to increase a chance of the students' intake of the input within the limited number of classes. I also decided to give written and oral instructions in the three tests in Japanese to avoid misunderstanding and to complete the tests within the given class time. For example, I supplied a Japanese translation for difficult English words given on the task sheets used in the conversation tasks, since the purpose of the tests was to examine their acquisition of target rules rather than their acquisition of vocabulary.

5.6.4 The preliminary questionnaire and its results

Immediately after the two preliminary tests, the students in the TG filled out the questionnaire (Appendix 11), which was intended to obtain general information about the students' background in English learning. In the questionnaire they answered: (a) their nationality and first language, (b) their year of study at each institution, (c) whether they had studied and/or lived in English-speaking countries, and (d) whether in the academic year of 2001/2002 they were taking other English courses focused on communication along with my instruction or whether in the past they had taken any. I will now briefly report its results.

All 48 students in the TG were Japanese and their first language was Japanese. As for their studying and/or living experience in English-speaking countries, none of the 18 nursing college students had any such experience, while 11 of the 30 university students did (3 students in Class A, eight students in Class

B). Their length of study in an English-speaking country ranged from three months to three years. Each student's length of stay is shown in Appendix 1.

As to the students' experience of taking English courses focusing on communication, none of the 18 nursing college students were taking any in the academic year of 2001/2002, when they took my instruction in the second semester. At the university, all 10 students in Class A were taking one English communication course given by an English teacher, and five of the 10 students were taking one more course given by either an American or a Japanese teacher. The 20 students in Class B were either in the third or fourth year of their study and had all taken at least one English course focusing on communication before, but for the academic year 2001/2002, when they joined my study, only eight of the 20 students were taking at least one course focusing on communication (1 course=6 students; 2 or 3 courses=2 students). It should be noted that the students reported that on these courses no focus was placed on any question forms[22]. The other 12 students were not taking any courses in that academic year.

The 12 volunteers in the CG did not fill in the questionnaire and therefore I have no data on their experience of studying or living abroad and the number of other English courses they were taking during my study.

Thus, although the students' background in L2 learning was diverse, I included all 60 students (TG=48, CG=12) in the analysis of my study, since their different backgrounds were not a central factor. This is because they had learnt English mainly through formal education in Japan and were homogeneous in terms of their first language, which allowed me to observe characteristics and any predisposition among Japanese learners in the acquisitional process of question forms. In the following section, I will describe the materials developed for the noticing-promotion instruction.

5.7 Materials development for the instruction

As discussed in Chapter 1, this study attempts to address the needs of learners obliged to study English in large classes, where immediate teacher feedback on individuals' mistakes is difficult to provide. It also aims to examine the potential and limitations of teachers' creating opportunities for noticing the gap between the target language and their interlanguage through dialogical interaction between learners. My pedagogical choice in exploring answers was the *noticing-promotion approach.* In this instruction the students were given explicit grammar teaching and the noticing tasks, in which I attempted to raise their awareness of the target Q-forms and to facilitate noticing the gap. While performing the task, the students were encouraged to pay attention to their output, both their own and their partner's, to notice the gap, and to take an active role both in assisting their partner by scaffolding replies to notice and self-correct mistakes.

I needed to carry out my main study at the two institutions within a given period of time and I established that four 90-minute classes were available at each institution for lessons. Then I designed a series of four treatment sessions, which were given once a week for four consecutive weeks, and developed teaching materials. I will first show the flow of the four sessions and then explain the rationale for the noticing-promotion approach.

5.7.1 Structure of the four treatment sessions

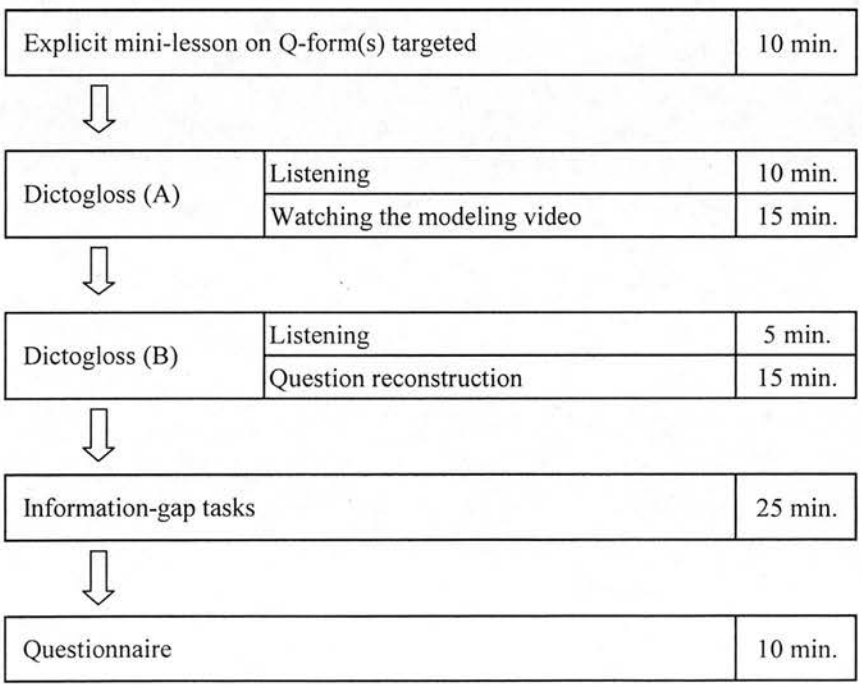
The structure of the four sessions was basically same, as shown in Figure 5.1 (see next page), although different Q-form(s) were targeted in the four treatment sessions. I will first briefly explain each component of the instruction, and then the target rules in each session in Section 5.7.3.

At the beginning of each session, I taught one or two Q-forms explicitly by using a handout for about 10 minutes in order to raise the students' awareness. Then I gave the student dyads dictogloss tasks (A) and (B). A dictogloss is a procedure that encourages learners to reflect on their own output; it consists of four parts: (1) preparation for dictation, (2) dictation - listening to a passage seeded with targeted L2

structures, (3) reconstruction of the passage in writing, while sharing notes from listening and working in pairs collaboratively, and (4) analysis and correction as learners reconstruct the passage. I used this procedure in order to promote gap-noticing between what the students knew about the target rule taught in the mini-lesson and what they could actually do with it in producing questions. They heard a recorded conversation between two native English speakers who were performing an information-gap task, which will be shown later, and asked to jot down the questions in the recording. The students then watched modelling on video showing how to carry out the reconstruction task in pairs from their own notes taken in listening. About 25 minutes were allocated for the first dictogloss.

In the second dictogloss, the students listened twice to a similar - but not identical - recorded conversation and reconstructed the questions which they had heard in the recording from their own notes taken in listening; however, the recorded conversation was at natural speed and the students did not have time to note down questions in their entirety; and they had to use their knowledge of the target grammar to reconstruct questions. The second dictogloss took about 20 minutes per session.

FIGURE 5.1 The flow of the noticing-promotion instruction



(Total 90 min./class)

Then the student dyads were given information-gap tasks. In the four treatment sessions different Q-forms were targeted and different tasks were prepared to maximise the effect of the instruction. The information-gap tasks used in each session were cognitively more demanding than the dictogloss reconstruction task, and more communicative pressure was put on the students. In each session, the same length of time (about 25 minutes) was allocated to carry out the information-gap tasks.

At the end of each session, about 10 minutes were spent on the questionnaire, in which the students were asked to rate the usefulness of some components of the instruction, such as the modelling shown on video, and to report their mistakes in using the target rule and their experience of giving and receiving assistance in pair work.

5.7.2 Rationale for the noticing-promotion approach

The rationale behind the noticing-promotion approach draws partly on the implications of information-processing models of cognitive theory for L2 learning, which were discussed in Chapter 3, in which the acquisition process is explained in terms of learners' limited capacity to process information. Findings from research into the acquisition of English question formation suggest that instruction can help learners to acquire structures if it focuses on those at the next stage of their development, namely, processable structures (e.g. Pienemann & Johnston, 1986; Pienemann, 1998). The noticing-promotion approach therefore targeted four Q-forms at stage 5 or stage 6, since many students seemed to have reached stage 4 or higher and were ready to learn them. Research findings on cognitive aspects of language learning also suggest the importance of raising learners' awareness of L2 features in order to increase the chance of noticing and intake of what is noticed into a developing interlanguage system (Schmidt, 1990). In the lessons, therefore, the target rules were explicitly taught at the beginning of each session to make them salient. In my experience, Japanese learners prefer an approach to language that draws upon concrete points taught in an explicit, rational way, and the introduction of the target rule in the mini-lesson seemed useful. Then the dictogloss and the modelling were used to promote further noticing and awareness-raising. Furthermore, in order to allow the students to use their limited attentional resources sparingly and effectively on the intake of knowledge of the target rule, the easier tasks (i.e. dictogloss) were given before the cognitively more demanding tasks (i.e. information-gap

tasks). A methodological procedure of presenting L2 features through explicit teaching prior to production has been criticised by advocates of the strong version of CLT, but it is still widely practised in EFL contexts and recognised as a useful means of raising learners' awareness within the limited period of class time. In designing the instruction, I considered changing the sequence of the components, but I rejected this idea because the dictogloss was intended to facilitate noticing through interaction with a partner, and in order to promote such noticing, it was necessary to raise the students' awareness of the target rules prior to this task.

Some may argue that the methodological procedure of the instruction is much like the traditional P-P-P. Indeed, the instruction began with explicit grammar teaching, followed by the easier tasks before the cognitively more demanding tasks, intended to lead the students from controlled to automatic use of new L2 structures and, as discussed earlier in Section 3.5.3, this teaching approach represents a compromise between FonF and FonF*S*, with a strong leaning towards the latter, in view of the inclusion of explicit grammar teaching at the beginning, and a narrow and intensive focus on one or two pre-selected structures in controlled activities, though not as far as FonF*S* as defined by Long (1991). The view underlying the noticing-promotion approach was different from that of the traditional P-P-P. In the noticing-promotion approach, the tasks were seen as a means by which the students acquire new knowledge and reconstruct their interlanguage through noticing the gap, not merely as a means by which they can activate their knowledge or improve fluency by producing L2 features. Seen in this light, this approach draws on a process view of L2 acquisition, where language learning is 'a series of systems, known as interlanguages, which are gradually grammaticised and restructured as learners incorporate new features' (Ellis, 2003:29). This is incompatible with the traditional P-P-P view of teaching a series of items which can gradually be accumulated.

The rationale behind the noticing-promotion approach also draws partly on the implications of interactive models of SLA theory, discussed in Chapter 2, which suggest that restructuring of interlanguage can be facilitated by learners' active involvement in the negotiation of meaning with a *focus on form* (e.g. Long, 1991). The tasks developed for the lessons were intended to facilitate that *negotiation of meaning* by encouraging the students to pay attention to output, both their own and their partner's, and to offer feedback on mistakes made by their partner.

The rationale also draws partly on research into the role of tasks in L2 teaching, discussed in Chapter 2. Research findings suggest that learner interaction is most likely facilitated in information-gap tasks, especially two-way tasks in which each student in a pair holds a part of information to be elicited by their partner (e.g. Long, 1980; Pica, Canagy and Falodun, 1993). In each session therefore, two-way information-gap tasks were given to the students. Study findings also show the necessity of getting learners to engage in meaning-focused language use (e.g. Skehan, 1998; Bygate, Skehan, & Swain., 2001) and I attempted to design information-gap tasks in each session to provide such occasions.

The rationale further draws partly on the interface position of cognitive theory which suggests that learners have cognitive processing abilities to integrate explicit knowledge (i.e. knowledge consciously learned in the formal instruction of form) and implicit knowledge (i.e. unconsciously acquired knowledge) into the developing interlanguage systems, and on Anderson's Skill Acquisition theory (1993, 1995) which is based on this position, both discussed in Chapter 2. The noticing-promotion approach combined explicit grammar teaching with the noticing tasks, in order to facilitate three stages of learning required for skill acquisition: (a) attainment of declarative knowledge, (b) proceduralization of knowledge, and (c) automatizing or fine-tuning of procedural knowledge.

Lastly, the rationale draws partly on the sociocultural theory of mind, which suggests that classroom learning encompasses not only input-out processing in the learner's head but also scaffolding through collaborative dialogue between two learners (e.g. Lantolf, 1996, 2000a, 2000b), as discussed in Chapter 4. The tasks used in the instruction were therefore intended to encourage the students to pay attention to their partner's output while performing the tasks, so that they could notice mistakes in their speech and take an active role in helping each other make self-corrections. The instruction also included the videotaped modelling performed by a learner dyad, intended to give the students ideas for effective ways to assist their partner.

In the next section, I will explain the target interrogatives focused on in each treatment session, before discussing the content of each treatment session in detail.

5.7.3 The target Q-forms in each treatment session

The rules targeted in each session are summarised in Table 5.6.

TABLE 5.6 The target Q-forms

Week 1	*Q-tag*
Week 2	*Cancel-inversion*
Week 3	*Cancel-inversion* and *Q-tag*
Week 4	*Y/N-Negative and Negative-Aux/Do-2nd*

In the first two sessions, one Q-form was targeted: *Q-tag* in Week 1 and *Cancel-inversion* in Week 2. In Week 3, these two rules were taught again. In Week 4, two negative Q-forms, *Y/N-Negative* and *Negative-Aux/Do-2nd*, were targeted. Originally, I meant to teach one target rule in each session, but I eventually decided to teach the four rules, as mentioned above. I made this change in order to make the amount of input in each session appropriate for the students, especially for weaker ones. For example, although I had noticed in piloting the materials and in marking the Grammaticality Judgment Test that the learners tended not to realise subject-pronoun agreement in tag questions when the subject is 'there', as in 'There is a library near the university, isn't it?'. I was aware of the need to provide the students with opportunities to produce tag questions with 'there', but I confined the subject of tag questions introduced in Week 1 to words other than 'there' and introduced it in Week 3, since the inclusion of too many subject words in one lesson would risk confusion.

Likewise, in Week 2, I did not include occasions to produce *Cancel-Inversion* questions with 'there' as a subject in the subordinate clause, but introduced it in Week 3. I had noticed in my teaching experience and in piloting earlier versions of the materials that many learners tended to forget to place the subject and the copula in the subordinate clause when its subject was 'there', as in 'Do you know how many students in your class?'. Another common mistake was the placement of an auxiliary verb between the Q-word and the subject of the subordinate clause, as in 'Could you tell me how many flatmates do you

have?', and there was the need to provide opportunities to use this rule with various combinations of a subject and a verb/copula. This decision was also motivated by the fact that it was difficult to seed occasions to use 'there' as a subject in one information-gap task. As a result, in Week 3, I again taught the two rules introduced in the first two sessions, and the word 'there' was introduced as a subject of tag questions and of the subordinate clause of *Cancel-Inversion* questions.

Due to these changes, I needed to consider whether or not to teach the other two rules in Week 4 (i.e. stage 5 *Negative-Aux/Do-2nd* and stage 6 *Y/N-Negative*). Taking into account both the merits and demerits of teaching two rules in one lesson, I eventually decided to include these two negative Q-forms together. One advantage was that these two rules could be taught effectively by highlighting differences between the use of each rule in terms of word order and the pragmatically appropriate use of each rule. In fact, as reported earlier, many students chose a wrong answer in Q26 in the Grammaticality Judgment Test, in which they were given an example for each rule (see page 95) and asked to judge whether the use of each rule was appropriate in inviting someone to do something together.

In both sentences, one of the two rules is correctly applied in terms of word order, but the use of *Y/N-Negative* question with 'Don't' in the first sentence is pragmatically inappropriate and the right answer here is 'only B is appropriate'; however, many students, even at higher developmental stages, answered that 'both questions are appropriate', 'only A is appropriate' or 'both questions are inappropriate', suggesting that they lacked knowledge of the pragmatically correct use of the two rules.

One disadvantage of my decision was that teaching the two rules in one lesson would reduce the amount of time spent on practising each rule, which might result in a less effective grasp of these rules. I was also aware of the possibility that teaching the two Q-forms categorised at two different developmental stages in one lesson might hinder the acquisition of these rules. However, I made this decision, since it seemed important to raise the students' awareness of the pragmatically appropriate use of questions - one of the neglected areas in L2 teaching.

I will now discuss, in turn, the content of the four treatment sessions.

5.7.4 WEEK 1

At the beginning of the first treatment session in each class, I explained that the purpose of the four treatment sessions was to help the students use a variety of question forms effectively in communication. I also told them that in each session they would be receiving a mini-lesson on one or two Q-forms which they seemed not to have acquired, and lots of listening and speaking tasks designed to provide opportunities to use the target rules.

I then told the class that they would be spending a lot of time in pairs and emphasised the importance of helping each other notice mistakes and make self-corrections. As explained in Section 5.4, I asked the students to find a partner by themselves prior to the pretest and the first treatment session. I was aware that learners' performance and the amount and quality of learner interaction in pair work could be affected by the partner's proficiency in the target language, such as developmental level and sex, but let them take the initiative. There were two reasons for this. Firstly, I intended to encourage the student pairs to scaffold each other at ease and expected that this was more likely when learners work with a partner of their own choice than with one chosen by the teacher. Secondly, my study was classroom-based and it was not possible to control various variables in pairing up the small number of students in each class.

5.7.4.1 Mini-lesson on *Q-tag*

In Week 1, the students were taught the use of *Q-tag*. Prior to the mini-lesson, I asked the class whether they had ever used tag questions for communication. Most students said that they knew the rule to some extent and had produced tag questions in writing and in decontextualised speaking exercises, but not for communication in and out of the classroom. I then told them that tag questions and other question forms targeted in the following three lessons were often used by native English speakers in daily conversations to enrich relationships through expression of ideas and feeling. Then I asked the students if they had heard someone producing tag questions in films and daily conversations. Most students answered that they had not, although a few university students who had studied in an English country answered in the affirmative. I then told the class that it is important for learners to

notice the gap between what they thought they knew about the rule and what they could actually do with it in communication, since it was the first step in L2 learning. I went on to say that once they took my lessons, they would become able to notice not only mistakes in producing tag questions by themselves and by peer learners, but also the use of tag questions in native speakers' speech, and that they would be surprised at why they were not used to noticing them before. This little talk was intended to raise the students' interest in the instruction and their awareness of the importance of *noticing the gap*.

I then gave the students a copy of the handout (Appendix 12) and explained the function and form of tag questions, focusing on the use of agreement and intonation. *Q-tag* is usually taught at senior secondary schools in Japan and learners are supposed to know the term 'tag question' in Japanese and its *form* - a tag is added on to a statement. However, they tend not to be given opportunities to practise this rule through communicative tasks and are not familiar with either its functions, such as checking whether something is true, or asking agreement. They are also not familiar with the appropriate use of intonation in tag questions.

In teaching agreement, I focused on (a) the use of negative tags after affirmative sentences, and vice versa, and on (b) subject-pronoun agreement, since I had observed many mistakes in these aspects while piloting the materials and examining the students' answers in the Grammaticality Judgment Test. For example, one common mistake was that learners put a negative tag after a sentence, even if the negative form of verb/auxiliary verb was used in the preceding sentences (e.g. 'Mike doesn't have a computer, doesn't he?'). Another common mistake was the use of the pronoun 'it' in the tag, irrespective of the subject of the main clause in the tag question (e.g. 'The woman is talking on the phone, isn't it?'). So in the handout I included listening exercises to raise the students' awareness of agreement (see Exercise A and B).

Regarding the use of intonation in tag questions, many learners seemed not to understand that rising intonation is used if the tag is a real question – if the speaker really wants to know something and is not sure about the answer - and that falling intonation is used if the tag is not a real question – if the speaker *is* sure of the answer. So I included some listening exercises and asked the students to assess the speaker's certainty about the statement in each tag question, by circling either 'sure' or 'not sure' printed

on the handout (see Exercise C and D).

5.7.4.2 First dictogloss task (W1)

After the mini-lesson, the paired students were given the first dictogloss. They heard a recorded dialogue in which two native English speakers (female and male) produced ten tag questions from the script provided, which will be shown later, while performing an information-gap task called Memory Game. This game was given to the class after the second dictogloss and the use of this dialogue had two functions: (a) to provide the input seeded with the target rule in the dictogloss and (b) to familiarise the students with procedures for the subsequent task (i.e. Memory Game).

I will now describe procedures for the first dictogloss task.

(1) Before listening to the recorded dialogue, each pair was provided with a sheet of paper on which four portraits were printed in colour (Appendix 13) and asked to memorise the characteristics of each portrait.

(2) The students were given one minute to do so, and then told to turn over the portrait sheet.

(3) Each student in pairs was given one of two task sheets, (A) or (B), both of which are shown in Appendix 14. On the left of each sheet, there were the following nine physical descriptions: hair style, colour of hair, colour of eyes, nose, moustache, beard, glasses, ribbon, and earrings. On each sheet, different five places were blank, which constituted the information gap. They were asked to fill out the blanks individually from memory.

(4) I told the class that they were to listen to the recorded dialogue between two native speakers playing the Memory Game with the same portrait sheet and task sheets.

The recorded dialogue included 10 tag questions and answers (see Appendix 15). A script had been used for the recording, to include various examples of question tags with both positive and negative forms of auxiliary verbs and with different pronouns (e.g. 'Holly has brown eyes, doesn't she?', 'Bill doesn't have a moustache, does he?", 'Holly isn't wearing glasses, is she?'). The texts used in the four treatment sessions contained not only the target forms but also discourse markers, linkers and fillers, such as 'Are you ready?', 'Well, then, let's start!', 'It's your turn.', 'I've won and you've lost!', 'Well/Umm…' and 'Let me see…', produced by two native speakers as they carried out different information-gap tasks, which will be shown later. The texts used in this study differed from other

dictogloss activities (e.g., Wajnryb, 1990; Swain, 1998) in that it used a spoken interactive text rather than a written one[23].

The students heard the recorded dialogue twice. Before the first listening, I told the class that the native speaker answering questions had a perfect memory of the portraits and I asked the students to check if their answers were correct while listening to the dialogue. In the second listening, I asked the students to take notes on the 10 tag questions and the responses on the listening sheet provided (Appendix 16), so that they could reconstruct the tag questions. On the sheet, there were ten spaces for notes for each question and the students were also asked to indicate the speaker's certainty about the statement in each tag question, as expressed by rising or falling intonation. The text used for dictation

After the second listening, I gave each pair a reconstruction sheet (Appendix 17) and asked them to reconstruct the tag questions together from their own notes. In the first dictogloss, however, instead of reconstructing the questions, the students were asked to watch the modelling on video to learn how to carry it out.

5.7.4.3 Modelling (W1)

In L2 classes, modelling is usually given by the teacher in person, but in this study I videotaped a Japanese female dyad performing the reconstruction task and used their performance as a model. This is because in the noticing-promotion approach the students were encouraged to work cooperatively, and showing good examples of pair work by the learners seemed more effective than giving modelling by myself. To the best of my knowledge, very few attempts have been made to examine the effect of modelling performed by learners, either in person or on video, on learners' performance in subsequent tasks, and I hoped that this study might give teachers ideas as to the potential and limitations of its classroom use.

Before explaining the modelling content, I will briefly introduce the two Japanese models shown on the video (Model A and B). They were foreign language teachers in their 30s and volunteered to help me to make the modelling video. Model A was teaching Japanese at the University of Edinburgh at that

time. Model B was an English teacher at a senior high school in Japan, who was visiting Edinburgh for a year to study English at IALS. In videorecording the modelling, they performed the reconstruction task from the script provided, which will be shown later, but were asked to pretend they were working spontaneously. While videotaping the modelling, each model had a copy of the script in front of them on the table, but on the videotape this sheet looked to the students like the listening notesheet which they had used. It took about two hours to videotape the modelling for each session.

The models acted out the script very well, but I was afraid that some parts of their conversations might look overdone in the students' eyes. So, before showing the modelling in the first session, I explained to the students that the learner pair shown on the video had been studying English in Edinburgh and taken a series of four English lessons, and that the way they had been working together in a pair was very efficient. I then added that I had asked them to perform the reconstruction task given in each lesson again after each class, so that I could videotape their performance in order to use the recording in my teaching in Japan. When the students heard my explanation, they looked interested, as I had expected, since it is rare for learners to have a chance to see peer learners performing the task.

Thus, the modelling video was intended to help the students understand how to do the reconstruction in a pair, but there was another purpose. I was interested in examining whether the inclusion of good examples of peer assistance could promote active interaction and noticing in the subsequent tasks. In modelling by a teacher alone or with a student, it is not possible to show examples of peer assistance for real, but in the modelling performed by the learner pair it is possible to seed this type of input. I prepared two kinds of dialogue scripts for modelling for each session and called them (a) the *peer-assistance version* and (b) the *self-correction version*.

In both versions, Model A reconstructed the first five tag questions (Q1-Q5) and Model B wrote down the reconstructed questions on the answer sheet provided (Appendix 17) [24]. Then, they changed roles and Model B reconstructed the other five tag questions (Q6-Q10). They worked cooperatively to reconstruct the same number of questions with the same Q-forms in both versions, but there were differences in some scenes in the way self-corrections were made. For example, in the two versions of the modelling video used in Week 1, in reconstructing 10 tag questions the student dyad made eight

mistakes in total (see Appendix 18 for the script of the *self-correction version*; see Appendix 19 for the script of the *peer-assistance* version) . In both versions, five of the eight mistakes involved agreement in tag questions (in Q2 and Q5 by Model A, Q6, Q8 and Q 10 by Model B), one mistake concerned intonation in a tag question (in Q1 by Model A), and two mistakes involved singular/plural distinction (in Q7 and Q9 by Model B). Two of the 10 reconstructed questions (Q3, Q4) included no mistakes.

I will now explain in detail the differences between the two versions used in Week 1. In the self-correction version all mistakes were self-corrected without peer assistance. In the peer-assistance version, six of the eight mistakes were self-corrected with the assistance from the partner, and the other two were self-corrected without assistance, as in the other version. In order to illustrate how the same mistakes were corrected with or without peer assistance, I will cite an example from each version, in both of which Model A made a mistake in agreement. In the extracts shown below, the utterances by Model B, who was writing down the reconstructed question, are in brackets.

Self-correction version:

> Question 2. Holly isn't wearing glasses, isn't she? ↑ (Holly isn't wearing glasses isn't she?) Uh, no, not 'isn't-isn't', 'isn't-is', negative-positive. (Yeah) Ok, then, Holly isn't wearing glasses, is she? ↑ (Holly isn't wearing glasses, is she? ↑ Is that ok?) Yes. Okay, next.

Peer-assistance version:

> Question 2. Holly isn't wearing glasses, isn't she? ↑ (Umm. Your intonation was correct, but you've said 'isn't-isn't', negative–negative.) Uh, yes, it should have been negative-positive. (Yeah) Ok, then, Holly isn't wearing glasses, is she? ↑ (Correct. Holly isn't wearing glasses, is she? ↑ Is that ok?) Yes. Okay, next.

In both versions, Model A used the negative form of the auxiliary in both the statement and the question tag, as in 'Holly isn't wearing glasses, isn't she?', and agreement was not realised. In the self-correction version the speaker noticed the mistake by herself without any signals from her partner, as in 'Uh, no, not 'isn't-isn't', 'isn't-is', negative–positive', and corrected the mistake. On the other hand, in the peer-assistance version Model A's partner noticed the mistake and gave Model A a hint, as in '(Umm. Your intonation was correct, but you've said 'isn't-isn't … negative–negative)', and this helped Model A to notice her mistake and to correct it.

Below is another pair of examples from each version, involving a mistake in intonation, which was the

other focus in the grammar teaching.

Self-correction version:

Question 1. The speaker said Holly has brown hair (Holly has brown hair), doesn't she? ↑ (Holly has brown hair, doesn't she? ↑ Um.. No, doesn't she? ↓ Falling intonation because the speaker on the tape was sure about the statement. (Okay, then, Holly has brown hair, doesn't she? ↓ Is that correct?) Yes. Okay, next.

Peer-assistance version:

Question 1. The speaker said Holly has brown hair (Holly has brown hair), doesn't she? ↑ (Holly has brown hair, doesn't she? ↑) Yeah. (Um.. doesn't she? ↑) Mhm .(Um…) Has-doesn't … (Mm..) Is it wrong? (Yeah… intonation.) Um…doesn't she? ↑ doesn't she? ↓ (Was the speaker on the tape sure about the statement?) Yes. (So…) Uh, yeah, the speaker used falling intonation! (Yes). Okay, then, Holly has brown hair, doesn't she? ↓ (Holly has brown hair, doesn't she? ↓ Is that correct?) Yes. Okay, next.

In both versions, Model A reconstructed a tag question with rising intonation, as in 'The speaker said Holly has brown hair, doesn't she? ↑ '. In the self-correction version, Model A checked her note of the speaker's certainty and noticed that she should have reconstructed this question with falling intonation, and then self-corrected the mistake while telling herself that 'Um.. No, doesn't she? ↓ Falling intonation because the speaker on the tape was sure about the statement'. In the peer-assistance version, Model A's partner noticed the same mistake made by her partner and repeated the question tag with rising intonation, as Model A had done. However, Model A did not notice what was wrong with her reconstructed question and asked Model B for help, 'Is it wrong?', and Model B gave her partner a hint by answering 'Yeah… intonation'. At this stage, Model A was still not sure of the problem and uttered the question tag twice with each intonation. Then, Model B voluntarily offered further help by asking 'Was the speaker on the tape sure about the statement?'. This helped Model A to notice that she should have paid more attention to intonation and make a self-correction.

Thus, assistance seeded in the peer-assistance version was provided in the following ways, singly or combined:

(a) repeating the partner's erroneous question utterance,

(b) giving a clue to help the partner find the area of mistake,

(c) giving the partner enough time to think, and

(d) providing an explanation about how the speaker's certainty of the statement is expressed with

intonation in the tag question.

Some were more direct than others, but all the assistance was given cooperatively and was appreciated by the person who had made the mistake. It should be noted that in the modelling, negative feedback in the form of recasts was not included as a means of assistance, because the purpose of the modelling video was not to encourage the students to correct their partner's mistakes but to encourage them to scaffold each other by assisting their partner to notice their mistakes, so that their partner could self-correct them. I hoped that the inclusion of good examples of various kinds of assistance seeded in the modelling would help the students to learn effective ways of offering assistance and would motivate them to try them out in the subsequent tasks. I also hoped that their frequent experience of giving and receiving peer assistance would increase their chances of not only noticing the gap but also acquiring the target rule.

At the university, Class A watched the self-correction version throughout the four treatment sessions and Class B watched the peer-assistance version. At the nursing college, there was only one class and I originally meant to use the peer-assistance version only, but the classroom was too small to make good-quality recordings of the students' conversations in the speaking tasks and I had to divide the class into two groups (Group A and Group B). So I had an assistant teacher, one of the examiners at the nursing college, and Group A was guided to another classroom and given the instruction by the assistant teacher, while Group B stayed in the same room and received instruction from me I prepared teacher's notes for the assistant for each session (see Appendix 20 for an example of teachers notes used in the first session), so that the two groups of students would receive the same instruction. Group A watched the self-correction version, whereas Group B watched the other. Before dividing the class, I told the students that the classroom was too small to record conversations at the quality necessary for the teacher to use them to improve her teaching. The intention was the students should not notice that the other group was shown a different modelling. The type of modelling video shown to each class/group of students at the two institutions and the number of students in each class/group is summarised in Table 5.7.

TABLE 5.7 Type of modelling video watched by each class/group of students

			Modelling video		Teacher
			self-correction version	peer-assistance version	
University	Class A	(N=10)	○		myself
	Class B	(N=20)		○	mysef
Nursing College	Group A	(N=8)	○		myself & assistant
	Group B	(N=9)		○	myself

It should be noted that, despite the differences between the two versions of modelling, both versions were seeded with the same English technical terms necessary to talk about the target rule, such as 'rising/falling intonation' and 'negative/positive', which were also included in the handout used in the mini-lesson. Added to these, other L2 words, such as 'singular/plural', were also seeded in both versions. Teachers tend to think that learners should know such basic words and that they can use them in giving assistance in the task. Learners, however, can hesitate to talk about grammar in the target language due to lack of confidence in using such terms and I seeded these words in order to facilitate the students' use of them when a need arose in the subsequent tasks. In addition, both versions were seeded with questions useful for the management of conversations, such as 'Are you ready?', 'Is that alright?', and 'Could you pass me the sheet?', in which an untargeted Q-form was applied.

5.7.4.4 Second dictogloss (W1)

After watching the assigned modelling video, the student dyads were given the second dictogloss. This time they simply listened to the recorded dialogue in which the same native speakers were carrying out the Memory Game with a different portrait sheet and task sheets. The dialogue consisted of 10 tag questions and the students listened to the dialogue twice. In the first listening, being told that the native speaker answering the questions had a perfect memory of the portraits, the students were asked to pay attention to the dialogue content without taking notes, so that they could learn how native speakers produce tag questions, such as agreement and intonation to express certainty of their statement. In the second listening they took notes on a listening sheet provided (Appendix 21) while listening to the dialogue, as in the first dictogloss, so that they could reconstruct the questions later by using their own

notes.

After the second listening, each pair was provided with a copy of a reconstruction sheet (Appendix 22) and one student reconstructed the first five tag questions and the other student wrote them down on the sheet. Halfway through, they changed roles and reconstructed all ten questions. They were also required to indicate the speaker's certainty in each tag question by choosing either 'sure' or 'not sure'. Before starting the task, I encouraged the student pairs to pay attention to their output and to cooperate with each other, as shown in the modelling. The students who were writing down questions reconstructed by their partner were further encouraged to help their partner notice mistakes in reconstructing tag questions, irrespective of the modelling content they had watched. Thus, the reconstruction task was intended to provide the students with opportunities to consolidate their knowledge of the target rule and to practise it through listening, speaking and writing with their focus on form.

In the reconstruction tasks and in the subsequent information gap tasks, the students were allowed to refer, where necessary, to the handout given in the mini-lesson. During the reconstruction task, the teacher walked around the classroom to observe the students' performance and took notes of their mistakes, but did not give individual feedback; the assistant teacher did the same. After reconstructing the ten questions, the teacher gave each pair a copy of the script of the recorded dialogue (Appendix 15) and asked the students in pairs to check (a) whether they were able to reconstruct the questions and (b) if their choice of the speaker's certainty for each tag question was correct. The teacher then gave the class summary feedback on common mistakes in using *Q-tag* in the task, but confined feedback to the focal grammar points in the mini-lesson to provide each class with the identical feedback (for the content of feedback, see the teacher's note in Appendix 20-3).

5.7.4.5 Information-gap task: Memory Game (W1)

After the second dictogloss, the students were given the Memory Game. As mentioned earlier, the information-gap task used in each session was cognitively more difficult than the reconstruction task. In the information-gap tasks used in the four sessions the students were required to focus on *meaning* in

order to elicit necessary information from their partner with their focus on *form*. In other words, the information-gap tasks were more demanding than the dictogloss reconstruction tasks.

For Week 1, I prepared two sets of portrait and task sheets for the Memory Game, (1) and (2). In order to give the reader a clearer idea about the content of this information-gap task, I will show the task sheets used in Memory Game (1) and explain its further procedures.

(1) Each pair was provided with a sheet of four portraits (Appendix 13) and asked to memorise the characteristics of each portrait in the given time (1 minute). After one minute, they were instructed to turn it over and keep it on their desk.

(2) Then, each student in pairs was given one of two task sheets, (A) or (B), on which they found nine blanks to describe the characteristics of the four portraits (Appendix 23). On the left of each sheet, nine items describing facial and other characteristics were printed. For instance, on the sheet (A), nine blanks involved hair style, colour of hair (twice), nose, moustache, beard, glasses, and a ribbon as a hair ornament (twice). The other columns on this sheet were shaded and the students holding the sheet (A) did not know which columns were blanked on their partner's sheet.

(4) After filling out the blanks individually from memory, the students holding the task sheet (A) were instructed to check their memory of the characteristics of each portrait by producing tag questions, as they had heard in listening in the dictogloss. In producing nine tag questions the students were instructed to pay attention to agreement, since the nine blanks were intended to elicit tag questions with a different subject word and a verb/auxiliary either in the affirmative or negative form. I summarise the nine tag questions expected to be produced by each student in pairs in Table 5.8 below.

TABLE 5.8 Model questions for the Memory Game (1)

Task sheet (A)			Task sheet (B)		
Mario	has black hair	, doesn't he?	Mario	has a small nose	, doesn't he?
Mario	doesn't have a mustache	, does he?	Mario	has a beard	, doesn't he?
Mario	isn't wearing glasses	, is he?	Mario	isn't wearing earrings	, is he?
Herman	is bald	, isn't he?	Herman	has red hair	, doesn't he?
Herman	doesn't have a beard	, does he?	Herman	doesn't have a mustache	, does he?
Herman	isn't wearing a ribbon	, is he?	Sally	is wearing glasses	, isn't she?
Sally	is wearing a ribbon	, isn't she?	Sally	isn't wearing earrings	, is she?
Paul	has white hair	, doesn't he?	Paul	has a mustache	, doesn't he?
Paul	has a small nose	, doesn't he?	Paul	doesn't have a beard	, does he?

(5) While the students holding the task sheet (A) were producing the nine questions, their partners needed to listen to the questions very carefully to answer, because they did not know which characteristic of which portrait their partner would ask about. They also did not have notes on those characteristics and needed to answer from memory. After producing one tag question for each blank, the students were allowed to ask more questions with other Q-forms. They changed roles and the students holding the sheet (B) asked nine questions. Before starting the game, I told

the class that their partner's answers might help them correct mistakes in their answers, but that their partner's answers might not be helpful all the time and they had to make a final decision by themselves.

(6) After each student in pairs produced nine tag questions, they were asked to turn over the portrait sheet provided at the beginning of the game to check if their final answers written in the nine blanks were correct. If their memory for the characteristics of the four portraits was correct, one point was given for each blank. On the task sheet, a space was provided to pencil in their score. This marking procedure was explained to the students before starting the game, in order to remind them that the task goal was to achieve a good mark by getting their answers right.

The students in pairs performed this game one more time with another portrait sheet (Appendix 13) and a set of task sheets (Appendix 24).

In this game, the students tried to memorise the characteristics of the four portraits within one minute, but it was not possible to recall all the nine characteristics blanked on their task sheet. Some characteristics were easier to recall than others, which was intended to provide occasions to produce a sentence in the form of tag questions with different levels of certainty and required them to use either rising or falling intonation. Thus, this game was designed to provide the students with lots of opportunities to practise tag questions and they were encouraged to pay attention to their output in order to notice their mistakes. However, it was predicted that the task demand (i.e. to ask questions from memory in order to get their answers correct) might induce the production of tag questions without attention to agreement and intonation. But this is a fact of learners' use of L2 in communication and it was hoped that the design feature of this game would encourage the students frequently to alter their attention from *meaning* to *form*, and vice versa, as need arose. In this way it was hoped to get them involved in linguistic and cognitive processes in producing questions with the target rule, which was supposed to provide good conditions for the acquisition of the target rule to take place.

5.7.4.6 Questionnaire (W1)

At the end of each treatment session, the same questionnaire was given to the students (Appendix 25). Its purpose was: (1) to elicit their comments on the usefulness of the instruction and (2) to encourage them to reflect both on their use of the target rule and on their efforts to notice mistakes in their output

and to give their partner assistance. There were 12 questions in the questionnaire. In the first five, the students were asked to rate the usefulness of each component of the instruction on a five-point scale from *1* (Not useful at all) to *5* (Very useful). For instance, in Q5, the students were asked if the modelling video was useful in helping them understand how to carry out the reconstruction in the dictogloss task. In the following question (Q6), the students were further asked to report on anything else they had learnt from watching the modelling video, apart from understanding how to do the reconstruction task.

The rest of the questions (Q7-Q12) were intended to encourage reflection on noticing mistakes and corrections, made with or without partner assistance. For example, in Q7 the students were asked to write down in spaces provided three kinds of mistakes which they had often made in the tasks. To the left of each space, two small boxes were printed next to each other and the students were asked to put a circle in the left-side box if they thought they had managed to correct the reported mistakes by themselves (Q8), and in the right-side box if they thought they had managed to correct with partner assistance (Q9). In addition, in Q10 the students reported three kinds of mistakes made by their partner, and then in Q11 they were asked to put a circle in a box if they thought they had managed to assist their partner to make corrections. In the last question (Q12), the students reported in detail the aspect of peer assistance which they had most appreciated. At the end of the questionnaire, a space was provided, so that the students could write anything which they had not been asked in the preceding part of the questionnaire.

5.7.5 WEEK 2

5.7.5.1 Mini-lesson on *Cancel-Inversion*

The target Q-form in Week 2 was *Cancel-Inversion*. I gave the students a copy of the handout (Appendix 26), on which I explained how to ask *yes/no* and *Q*-word questions politely using this rule. Giving the three examples of direct *yes/no*-questions on the handout, I showed how the same thing could be asked more politely with the target rule by using the first example ('Do you have a spare room?'), as in 'Could you tell me if you have a spare room?'. I explained that in this question the main clause 'Could you tell me' was followed by a clause in which inversion was not present with a word 'if' placed at the beginning. Then, I asked the students to compose polite questions in writing from the other two examples.

After giving the correct answers to the two questions, I moved on to teach how *Q*-word questions could be asked in the polite form. After explaining that the main clause 'Could you tell me' was followed by a non-inversion sentence with the Q-word at the beginning of the subordinate clause, I asked the students to check accuracy of the given example, in which the auxiliary verb 'does' was placed between the *Q*-word and the subject in the subordinate clause ('Could you tell me what colour of eyes does Anita have?'). In each class at the two institutions some students found the mistake and I wrote down the correct question on the whiteboard in front of the classroom, as in 'Could you tell me what colour of eyes Anita has?', drawing their attention to number agreement between the subject and the verb in the subordinate clause. Then, I asked the class to frame polite questions orally by using the other two *Q*-word questions. I told the class that learners could make mistakes in speaking more frequently when producing indirect questions than in writing, emphasizing the importance of paying attention to their output in the tasks to follow. In addition, I introduced another expression for the main clause 'Do you know', as a variation to form indirect questions, and wrote it on the whiteboard.

5.7.5.2 First dictogloss (W2)

The mini-lesson was followed by the first dictogloss, as in Week 1. The students heard a recorded

dialogue between two native speakers, performing an information-gap task called 'GUESS WHO?' (Hasbro International Inc.) from a script (Appendix 27). Although this is a popular board game among children in English-speaking countries, none of the students in my study had played it before. Before listening to the recorded dialogue, each student in pairs was provided with a set of game materials, used by the native speakers in the recording, so that they could understand the context in which the recorded conversation was taking place. This was also intended to help them understand the procedure of the game, which was given after the second dictogloss as an information-gap task. I will explain the procedure by describing what the students were asked to do before listening.

(1) Each student was given a set of 24 different colour portraits (approximately 2.5×3.4 cm), including six female and 18 male (Appendix 28). On each card was printed the name of the person shown on each portrait. Some of the characters on the portraits share features, such as a beard, glasses, a hat, or baldness.

(2) Each student in pairs then placed their set of 24 portraits face-up in front of them[25].

(3) Each student drew one portrait from another set of the same 24 portraits printed on larger size cards (3×5cm) and kept it secret from their partner during the task by placing it in the card-holder provided. This became the mystery person for their partner. In the first dictogloss, the students facing the corridor outside the classroom were given Anita's portrait, held by one of the two native speakers (Mary). Those facing the other side of the classroom were given Paul's portrait, held by Mary's partner (Tom).

(4) Then I told the students that they were to listen to the dialogue in which the two native speakers were playing this game with the same mystery persons. I explained that in this game the players took it in turn to ask questions one at a time, to identify which portrait their partner was holding.

(5) I played the recorded dialogue and stopped it after the first question 'Is your person bald?' produced by Mary and the answer to that question 'No' produced by Tom. I then asked the students holding Anita's portrait in their card-holder to discard the unnecessary portraits and to keep the discarded ones aside on the desk, as Mary was supposed to be doing in the recording.

(6) Then I played the next recorded question and answer and asked the students holding Paul's portrait to discard unnecessary cards. When discarding cards, learners can get confused. So, for practice, I instructed the students to keep playing the game along with the recording.

(7) In the recording, Tom came to know his mystery person first and asked 'Is your person Anita?' and won the game[26].

After making sure that the students understood the procedures for this game, I provided each student with a listening sheet (Appendix 29) and asked them to note down the eight questions on the sheet while listening once to the recording, so that they could reconstruct them from their own notes. The eight questions consisted of two *Cancel-Inversion* questions, five *Y/N-Inversion* questions, and one

Affirmative-Aux/Do-2nd question. After the second listening, however, the students did not reconstruct the questions; instead, they watched one of the two versions of modelling video, assigned as in Week 1.

5.7.5.3 Modelling (W2)

In both versions of modelling, Model A reconstructed the first four questions, while Model B wrote down the reconstructed questions on the sheet provided. Then they changed roles. Four of the eight questions were reconstructed without mistakes (Q1, Q5, Q7, Q8), but in producing the other four questions (Q2, Q3, Q4, Q6) the same model made the same mistakes in reconstructing two *Cancel-Inversion* questions (Q3, Q6), one *Y/N-Inversion* (Q2), and one *Affirmative-Aux/Do-2nd* question (Q4). In the self-correction version the speaker noticed her mistakes and self-corrected them without partner assistance, while in the peer-assistance version the speaker's partner noticed the mistakes and helped the speaker to make self-corrections. To illustrate this, I will quote an example from each version (Q3), in which Model A placed an inverted sentence in the subordinate clause, as in 'Could you tell me if .. is your person's hair white?'. The full scripts for each version are shown in Appendix 30 (self-correction version) and in Appendix 31 (peer-assistance version).

Self-correction version:

> Question 3. Could you tell me (Could you tell me) Could you tell me if .. is your person's hair white? (if is your person's hair white?) Um.. No. My word order isn't correct. (Word order?) Yeah. After could you tell me, there comes a non-inversion sentence. (Um...What does non-inversion sentence mean in Japanese?) *Tochi-nashi-bun* (*Tochi-nashi-bun* Uh, yes! *Tochi-nashi-bun*, non-inversion sentence!) Okay, then. Could you tell me if . your person's hair is white? (if your person's hair is white? Could you tell me if your person's hair is white? Is that ok?) Yes. Okay, next.

Peer-assistance version:

> Question 3. Could you tell me (Could you tell me) Could you tell me if .. is your person's hair white? (Umm..I'm afraid your word order is not correct) Word order? (Yes) Mm...(Okay, after could you tell me, there comes a non-inversion sentence.) Non-inversion sentence? (Yes) Um...What does non-inversion sentence mean in Japanese? (*Tochi-nashi-bun*) *Tochi-nashi-bun* Uh, yes! *Tochi-nashi-bun*, non-inversion sentence! (Yes, that's right) Okay, then. Could you tell me if . your person's hair is white? (if your person's hair is white? Could you tell me if your person's hair is white? Is that ok?) Yes. Okay, next.

In the self-correction version, Model A noticed her mistake and said 'Um.. No. My word order isn't correct. After could you tell me, there comes a non-inversion sentence'. Her partner was not sure

what 'non-inversion sentence' meant and asked Model A its meaning, and then Model A translated it into Japanese, as in 'Tochi-nashi-bun'. Model A then produced a *Cancel-Inversion* question successfully.

In the peer-assistance version, Model A did not notice her mistake, so her partner pointed it out to her, as in 'Umm.. I'm afraid your word order is not correct'. However, Model A still did not identify the problem, so her partner added, 'Okay, after *could you tell me*, there comes a non-inversion sentence.' Model A did not know the meaning of the phrase 'non-inversion sentence' and asked her partner its meaning, and then Model B translated it into Japanese. This helped Model A to remember the rule and to self-correct the mistake.

In both versions, I included the question about the meaning of 'non-inversion sentence', as in 'What does non-inversion sentence mean?', hoping that the inclusion of that interaction would facilitate metatalk in pairs and remind them that they were allowed to ask meaning of technical terms used by their partner in the subsequent tasks.

5.7.5.4 Second dictogloss (W2)

A similar but not identical recorded dialogue was prepared for the second dictogloss, in which the same two native speakers played the same game (GUESS WHO?) from the script (Appendix 27) with a different portrait card as their mystery person. They produced ten questions, including three *Cancel-Inversion* questions, five *Y/N-Inversion* questions, and two *Affirmative-Aux/Do-2nd* questions. As in the first Dictogloss, the students in pairs played the game along with the recording and in the second listening, they took notes on a listening sheet (Appendix 32). In the reconstruction task, one student in each pair reconstructed the first five questions and the other student wrote down the questions on the sheet provided (Appendix 33). Before starting this task, they were reminded to pay attention to their output in order to notice mistakes and to cooperate with each other as shown in the modelling video. The same instructions were given in every treatment session before they did the reconstruction task in the second dictogloss. As in Week 1, the teacher walked around the classroom and observed the students' performance and took notes on their mistakes. After reconstructing the ten questions, each

pair was given a sheet of on which they found a script of the recorded dialogue and checked their answers. The teacher gave summary feedback on common mistakes, but confined feedback to the points focused on in the mini-lesson, as in Week 1. Then, the teacher asked the class if they had paid attention to their output in order to notice mistakes and if they could have been able to help their partner to make self-corrections. The students were then encouraged to keep cooperating with each other in the next task.

5.7.5.5 Information-gap task: GUESS WHO? (W2)

The students in pairs played the information-gap task, 'GUESS WHO?', twice. They were encouraged to produce at least three indirect questions using the target rule (*Cancel-Inversion*). In the first match, the students were allowed to ask only *yes/no*-questions. In the second play, in addition to *yes/no* questions, they were allowed to ask *wh/Q*-word questions but only in indirect questions. If they finished the match earlier than the other pairs, they were allowed to carry on playing the game. Thus, the game had a competitive element and the students' attention seemed to be drawn largely to producing questions.

5.7.5.6 Questionnaire (W2)

The students filled in the same questionnaire used in Week 1.

5.7.6 WEEK 3

5.7.6.1 Mini-lesson on *Cancel-Inversion* and *Q-tag*

In Week 3 the two rules taught in the first two sessions (*Q-tag* and *Cancel-Inversion*) were targeted again. As mentioned earlier, the word 'there' was excluded as a subject in tag questions in Week 1 and as a subject in the subordinate clause of *Cancel-Inversion* question in Week 2, but in Week 3 all examples given on the mini-lesson handout on involved the word 'there'.

Cancel-Inversion

I first focused on the use of *Cancel-Inversion* and pointed out two common mistakes in producing indirect questions when the word 'there' was used as an introductory subject in questions:

(a) missing subject and copula in the subordinate clause, and
(b) presence of inversion in the subordinate clause.

I gave three question examples on the handout (Appendix 34) and asked the students to find the mistakes and correct them. For the first example 'Could you tell me how many people in the room?', the subject and the copula ('there are') were missing in the subordinate clause. This mistake was pointed out by students in the two university classes, while in the class at the nursing college I had to explain the mistake.

The second example was a direct *Q*-word question ('How many students in your class?'), in which the same missing phenomenon was seeded. This mistake was pointed out by students in all classes. I told them that learners could make this mistake even in such questions, especially in speaking, and I pointed out the need to concentrate when noticing mistakes in the speaking tasks given later in this session[27].

In the third example 'Do you know what kind of things are there in that cupboard?', inversion was present in the subordinate clause; this mistake was found by students in each class. I told the class that this was a common mistake in producing indirect questions with the *Q*-word when the subject of the

subordinate clause was 'there'. I then pointed out that its cause was learners' tendency to produce the copula and the subject as a chunk in questions, as in 'are there', without paying enough attention to output.

Q-tag

I then moved on to teaching the use of *Q-tag* and asked the students to correct mistakes on the handout in two tag questions, in which 'there' was used as a subject and subject-pronoun agreement was not realised, as shown below.

*In your picture, there is a volleyball under the table, isn't it?
*In your picture, there are three trees, aren't they?

The mistake in each question was pointed out by students in each class. I asked them if they had produced tag questions with 'there' as a subject of the sentence. Most students answered 'no'. I then told the class that they would first be given two tasks to practise *Cancel-Inversion*, and then one task to practise *Q-tag*. I encouraged them to pay attention to their output.

5.7.6.2 First dictogloss (W3)

In the recorded dialogue used in the first dictogloss, there were 13 questions, including one *Affirmative-Aux/Do-2nd* question with 'there' as a subject of the sentence, seven *Y/N-Inversion* questions and five *Cancel-Inversion* questions (see Appendix 35 for the script). Of the five *Cancel-Inversion* question examples, in three questions 'there' was used as subject of the subordinate clause (e.g., 'Could you tell me how many people there are in your picture?') and in the other two an auxiliary 'is' was used in the subordinate clause (e.g., 'Could you tell me what the man is doing?'). In this dialogue, no *Q-tag* example was included to help the students concentrate on one target rule at a time. In the recording, the same two native speakers were performing an information-gap task called 'Our Neighbours'.

Before listening to the recorded dialogue, each student was given one of two task sheets, (A) or (B), (see Appendices 36 and 37, respectively), used by each native speaker in the recording. The student dyads

would be given this task later, and the instructions for the dictogloss were intended not only to explain the context in which the recorded conversation was taking place but also to help them understand the procedures for the information-gap task, as in the previous sessions. I will now explain the procedures for the information-gap task ('Our neighbour') below by describing what the students were asked to do before listening.

(1) Each student was given a copy of the aforementioned task sheets, either (A) or (B), on which they could see the actions of three people through the windows. In each flat four differences were seeded between the two task sheets.

(2) The students holding task sheet (A) were asked to find the differences for Flat 2, while those holding task sheet (B) were asked to find differences for Flat 1. In order to save time, the other two flats shown on both sheets were blanked out in the dictogloss and the students were told so.

(3) Before starting the task, the students were encouraged to produce *Cancel-Inversion* questions in finding differences. They were reminded that there were lots of opportunities to ask questions with 'there' as a subject, and that they had to pay attention to the common mistakes explained in the mini-lesson.

Then the student dyads heard the recorded dialogue twice. During the first listening, they noted the differences which the speaker was supposed to have found. In the second listening, they were asked to note the 13 questions produced in the recording on their task sheet, so that they could reconstruct the questions from their notes.

5.7.6.3 Modelling (W3)

Instead of reconstructing the questions after the second listening, the students watched one of the two versions of the modelling video, assigned as in the first two sessions. In both versions, the models reconstructed 13 questions, including five *Cancel-Inversion* questions, six *Y/N-Inversion* questions, and two *Wh-Inversion* questions. The dialogue script is shown in Appendix 38 (self-correction version) and in Appendix 39 (peer-assistance version). In both versions the same model made the same mistakes in producing one *Y/N-Inversion* question (Q1) and four *Cancel-Inversion* questions (Q3, Q9, Q12, Q13). As to one *Y/N-Inversion* question, the mistake involved the missing subject and copula, both pointed out in the grammar lesson. For these three *Cancel-Inversion* questions, the mistake involved the presence of inversion in the subordinate clause[28]. I will illustrate this below with one example for each version of modelling, in which Model B inverted the subject and the copula in the

subordinate clause, as in 'Could you tell me how many cushions are there on the sofa?'. In the self-correction version, Model B noticed her mistake and corrected it, while in the other version, Model B did not notice it, so her partner's prompt 'Really?' helped Model B to make a self-correction.

Self-correction version
Question 13. Could you tell me (Could you tell me) how many cushions are there (how many cushions are there) on the sofa? (on the sofa. Could you tell me how many cushions are there on the sofa? Is that right?) No! Non-inversion. Could you tell me how many cushions there are on the sofa? Yeah! That's right! We've finished.

Peer-assistance version
Question 13. Could you tell me (Could you tell me) how many cushions are there (how many cushions are there) on the sofa? (on the sofa. Could you tell me how many cushions are there on the sofa? Is that right?) Yes. (Really?) . Mmm.. No! Non-inversion. Could you tell me how many cushions there are on the sofa? Yeah! That's right! We've finished.

5.7.6.4 Second dictogloss (W3)

After watching the modelling, the student pairs were given the second dictogloss. They were given a similar but not identical task sheet, (A) or (B), shown in Appendices 40 and 41, respectively. They heard a recorded dialogue twice, which consisted of 12 questions and answers, including six *Cancel-Inversion* questions, three *Affirmative-Aux/Do-2nd* questions, one *Wh-Inversion* question and two *Y/N-Inversion* questions (see Appendix 35 for the scripts). In the first listening, the students noted differences which the speaker was supposed to have found on their task sheet. In the second listening, they took notes of all the questions on their task sheet. Then each pair was provided with a reconstruction sheet (Appendix 42). The students holding the task sheet (A) reconstructed the first six questions and the other student the remaining six questions, in both of which three examples of the target rule (*Cancel-Inversion*) were included. Before starting the reconstruction task, the same instructions given in the first two treatment sessions were repeated.

5.7.6.5 Information-gap task: Our Neighbour (W3)

After the second dictogloss, the student dyads performed the information-gap task called 'Our Neighbour', intended to provide the students with opportunities to practise *Cancel-inversion* question with the word 'there' as an introductory subject in the subordinate clause. The two task sheets used in

this game are shown in Appendices 43 and 44. In this task, the students holding Sheet (A) asked questions to find the differences in Flats 2 and 4, while those holding Sheet (B) asked questions to find the differences in Flats 1 and 3. Further procedures for this task were explained in 5.7.6.2. The pairs performed this task only once, since another information-gap task to practise the other target rule (*Q-tag*) was prepared for Week 3, which will be shown in the next section.

5.7.6.6 Information-gap task: Do you have a good memory? (W3)

I prepared an information-gap task called 'Do you have a good memory?', to provide opportunities to practise *Q-tag* with 'there' as a subject of the sentence. This game was similar to the information-gap task given in Week 1 (Memory Game):

(1) Each pair was given a task sheet (Appendix 45), on which they saw four flats. In each, one or two people were doing something. This sheet was called the 'teacher's picture'

(2) The students were asked to memorise the teacher's picture, so that they could recall the difference between the teacher's pictures and their own pictures given later. Four differences were seeded in each flat between the teacher's picture and the student's. They were given one minute and then asked to turn over the sheet.

(3) Each student in pairs was given one of two task sheets, (A) or (B), on which they saw similar, but not identical pictures shown in Appendices 46 and 47, respectively. They were asked to take notes of differences.

(4) The students holding sheet (A) were instructed to check if differences noticed in Flats 2 and 4 were correct by producing tag questions starting with 'In the teacher's picture'. Those holding sheet (B) were asked to check if differences found in Flats 1 and 3 were correct.

(5) Before asking questions, the students were given two minutes to indicate their certainty of their memory about each difference. If they *were* sure, they marked their notes with a '★'; if not, they did not put any mark. This procedure was intended to remind the students of the use of intonation in tag questions, which had been taught in Week 1.

(6) The students were then encouraged to pay attention to agreement, especially between the subject of the sentence and the pronoun in the tag. They were also encouraged to cooperate.

(7) At the end of the task, each student pair were asked to check how many differences they correctly spotted. They took a look at the teacher's picture sheet . Each difference correctly identified was awarded one mark and the students were instructed to put their own score in the column provided on their task sheet. This marking system was explained before starting the task.

5.7.6.7 Questionnaire (W3)

The students filled out the same questionnaire used in the previous two sessions.

5.7.7 WEEK 4

5.7.7.1 Mini-lesson on *Negative-Aux/Do-2nd* and *Y/N-Negative*

The target rules in the last treatment session were two negative Q-forms: stage 5 *Negative-Aux/Do-2nd* and stage 6 *Y/N-Negative*. In teaching, I called each rule *Wh-negative questions starting with 'why'* and *Yes/no negative questions*. I first focused on *wh*-negative questions negative questions.

Wh-negative questions

I introduced, in turn, two functions of *Wh-negative questions starting with 'why'*: (a) to ask reasons and (b) to press invitations and offers. Giving two examples of the application of this rule on the handout (Appendix 48), I asked the class if there were any mistakes in the two questions. In these examples the negative form of auxiliary was left after the subject in each, as in 'Why we can't use this chopping board?' and 'Why you haven't done shopping yet?'. The mistake was spotted by students in each class. I then explained that one had to place the negative form of the auxiliary before the subject. I told the class that misplacement of the negative form of the auxiliary was a common mistake in producing negative questions with the interrogative 'why'. I also pointed out that even at an advanced level learners could keep making this mistake without noticing, emphasizing the importance of being aware of the problem and the necessity of their conscious effort to pay attention to their output while carrying out the tasks to practise this structure later in the class. By way of summary, I said that one of the functions of negative questions with 'why' was to ask for reasons.

I then introduced another function of *Wh-negative questions with 'why'* (i.e. to press invitations and offers), using the example on the handout ('Why don't you come in for a few minutes?)[29]. I asked the student dyads to use this rule to press invitations and offers by having a little conversation with their partner. They had to pay attention to the word order.

Then I moved to the other target rule, *Yes/no-negative questions*. I explained that *Yes/no-negative questions* like 'Haven't you done the shopping yet?' were produced by native speakers to express surprise that something had not happened or had not been happening, as shown on the handout. I also

explained that *yes/no* negative questions were formed by placing the negative form of auxiliary before the subject.

I did not introduce other functions of *Y/N-Inversion* in the class, but to raise the students' further awareness of the appropriate use of this Q-form, I gave one example of *yes/no* negative question 'Can't you lend me your pen for a moment?' on the handout. In this question, although the *word order* is correctly applied, the use of the rule when asking someone to do something is pragmatically inappropriate. Then I asked the class to find a mistake in the example. The mistake was identified by students in both university classes, but not by those in the class at the nursing college. I then explained in each class that this *yes/no negative question* might be understood as a complaint or criticism, meaning something like 'Are you too selfish to lend me…?", and that native speakers usually use the ordinary question 'Excuse me, could you lend me your pen?' when asking people to do something.

5.7.7.2 First dictogloss (W4)

In the first dictogloss, the students heard a recorded dialogue between the same two native speakers, who were performing a guided role-play called 'I have a favour' from the script (Appendix 49). In this guided role-play, the two native speakers were Mary and Tom, a married couple. They were having a conversation in their house first thing in the morning, with Mary asking Tom to vacuum the living room, as in 'Would it be all right for you to vacuum the living room?'. Tom accepted the request willingly and Mary dashed off to work. Later that day, Mary came home and noticed that Tom had not vacuumed the room yet, which prompted her to pose a *Y/N-Negative* question, as in 'Tom, haven't you vacuumed the room, yet?'. Tom apologised and vacuumed the room right away. After that, Tom suggested Mary go skiing that weekend, but she said 'No', which provided an occasion to ask 'Why can't we go skiing?'. Then Mary answered that 'We have to save money. We're going to Japan next summer, aren't we?'. Thus, in this recorded dialogue one example was seeded for four Q-forms: *Y/N-Inversion*, *Y/N-Negative*, *Wh-Negative starting with 'why'*, and *Q-tag*. Compared with the recorded dialogues used in the dictogloss in the first three sessions, there was a smaller number of examples of each target rule included in this dialogue.

Before listening, each student was given a task sheet (Appendix 50), on which they saw the script of this recorded dialogue in a flowchart with the instruction for the guided role-play. In the chart, eight places were left blank and in order to fill in the blanks the students were asked to take notes of the dialogue while listening twice to the recorded dialogue.

5.7.7.3 Modelling (W4)

As in the first three treatment sessions, after listening twice to the recorded dialogue, the students watched the assigned version of the modelling video, instead of doing the reconstruction task. In both versions, the four questions were reconstructed by the same model from the script (see Appendix 51 for the self-correction version and Appendix 52 for the peer-assistance version). In the reconstruction of three of the four questions, the same model made the same mistake involving one of the following areas: (a) the pragmatically inappropriate use of *Y/N-Inversion* question in the context where a *Yes/no negative question* was supposed to be reconstructed, (b) misplacement of the negative form of auxiliary in a *Wh-negative question with 'why'*, and (c) failure in realising agreement in the tag question.

To illustrate how the same mistakes were corrected in each version of modelling, I will cite an example from each. In the two examples, the same model (A) produced a *Y/N-Inversion* question 'Have you vacuumed the room?' in the context of producing a *Yes/no negative question* to express surprise that Tom had not vacuumed the room. In the self-correction version, Model A noticed and corrected her mistake, while telling herself that she should have asked a negative question. In the peer-assistance version, Model A's partner noticed this mistake and repeated the erroneous question twice, as in 'Have you vacuumed?' and 'Have you?', but Model A did not notice. Then her partner tried to help Model A see that she should have expressed her surprise that Tom had not vacuumed the room and asked 'Did Mary ask if Tom had vacuumed the room?'. With this help, Model A noticed her mistake and corrected it successfully.

Self-correction version:

> Question four. Mary said 'Have you vacuumed (Have you vacuumed) Have you vacuumed . mm . Mary wasn't asking if Tom has vacuumed the room or not. (No) but she was surprised (Yeah) because she found Tom hadn't vacuumed the room yet. (Yeah) Okay, then, haven't . haven't you, negative question. (That's it) Haven't you vacuumed (Haven't you vacuumed) the room yet? (the

room yet? Haven't you vacuumed the room yet? Is that okay?) Yes. Okay, your turn.

Peer-assistance version:

Question four. Mary said 'Have you vacuumed (Have you vacuumed? ↑) mm.. Have you vacuumed (Have you?) ↑ mm.. (Did Mary ask if Tom has vacuumed the room or not?) No. (Then, why did Mary say this?) Because she was surprised when she found Tom hadn't vacuumed the room yet. (Yeah, and so..) mm . Uh! haven't . haven't you! (Yes) Negative question. (That's it!) Okay, then, haven't you vacuumed (Haven't you vacuumed) the room yet? (the room yet? Haven't you vacuumed the room yet? Is that okay?) Yes. Okay, your turn.

It should be noted that in Week 4, after watching the modelling video, the students were not given the second dictogloss. Instead, they worked on three kinds of guided role-plays, which will be shown later. I had to make this change because it was difficult to seed many occasions to use these two negative Q-forms in one task, and there was a need to provide the students with opportunities to practise these rules.

5.7.7.4 Three kinds of guided role-plays (W4)

The first guided role-play was intended to provide the students with opportunities to practise *Y/N-Negative* (Role-play 1) and the second to practise *Negative-Aux/Do-2nd* (Role-play 2). In the third role-play, the students practised both Q-forms (Role-play 3). I will now explain in detail each role-play in turn.

Role-play 1: *Haven't you…?*

In the first role-play the students in pairs practised how to use *Y/N-Negative* to express surprise that something had not happened. Before starting the task, I provided each student with a copy of the script used in the dictogloss (Appendix 49) and went through the dialogue line by line to teach the students the useful expressions seeded in it, so that they could use them in the guided role-play. For example, in the dialogue Mary started her request to vacuum the living room with 'Listen' in order to capture Tom's attention. Mary also used the expression unfamiliar to many students, 'I have to dash off', as she was leaving the house. With the two seeded negative questions, I drew the students' attention to their word order and their function in each question. *Q-tag* was not targeted in this session, but I explained its usage in the dialogue for revision.

Then each student in pairs was given a copy of task sheet, (A) or (B), (see Appendices 53 and 54), on which the instructions for another guided role-play were given in a flowchart. The content of the role-play was the same as in the dictogloss, but no lines were shown in the chart and the students had to perform the task by reading the instructions. In this role-play, each student was also provided with one of two sets of different four A5-size cartoon cards, (A) or (B), which are shown in Appendix 55 (Set A) and in Appendix 56 (Set B). Each set of double-sided cards was chained together with a small ring on top. On the front of each, the students saw a woman asking someone to do something in the morning. For example, the students holding the card set (A) saw on the front side of their first card a woman asking a man to paint a door in the morning. On the other side, they saw a picture in which later that night the same woman was surprised to find that her request had not been carried out. This provided an occasion to produce a *Yes/no negative question*, 'Haven't you painted the door yet?'.

The cards containing the requests to be made are summarised in Table 5.9 below. At the end of each set, an extra card was provided with a blank speech bubble, in order to encourage the students to make an unprompted request. The students were instructed to use their own names to address each other throughout the task.

TABLE 5.9 Cartoon cards used in Role-play 1

Set (A)	Set (B)
to paint the back door	to do grocery shopping
to mow grass in the garden	to tidy up the dining table
to hang out the washing	to empty the garbage bin
to vacuum the livingroom	to do the washing-up
[blank]	[blank]

Role-play 2: *Why can't we…?*

The second role-play was intended to give the students lots of opportunities to produce *Wh-negative questions with 'why'*. The students in pairs were provided with a task sheet (Appendix 57), on which they saw a flowchart for a dialogue between a man and a woman with four speech bubbles in which an incomplete sentence was given in Japanese. In the first speech bubble, the man was suggesting to the woman that they do something together, but the content of his suggestion was left blank, as in '... *syou*' (= '…let's)', with a choice of four time words '*Ashita/konshumatsu/kono fuyu* (*natsu*)' (= 'tomorrow/this weekend/this winter (summer)' at the beginning of the sentence. On the right of this speech bubble, in order to show the students several ways to ask someone to do something, three expressions were provided, as in 'Let's..../Shall we...?/Why don't we..?'. In the second speech bubble, they saw a reply from the woman in Japanese, as in '… *dekinaiwa* (= ….. can't)'. The third speech bubble had a question made by the man, as in '*Doushite … dekinaindai* (= Why ….. can't)?', which produced an obligatory occasion to use *Wh-negative questions with 'why'*. In the last speech bubble, the students saw an unfinished reply from the woman, as in '*Nazenara* … (= because …..)'.

On the right-hand side of the task sheet, I gave 15 sample reasons to help the students vary their replies to suggestions from their partner. This was also intended to provide L2 input. The reasons were written in Japanese with an English translation for each, but the students were encouraged by folding the task sheet or by hiding the English translation not to look at the English translation. With the lower-level students at the nursing college, the inclusion of the English translation was intended to help them carry out the task without falling into silence.

In addition to the task sheet, each student received one of two sets of 10 picture cards, (A) or (B), on which they found a man making a suggestion to a woman (see Appendix 58 for Set A and Appendix 59 for Set B). On the back of each card, a useful expression for making a suggestion was printed, but the students were encouraged to make suggestions without referring to them. At the end of each set, an extra card was provided with a blank speech bubble in order to help the students make suggestions without prompting.

Role-play 3: *Haven't you..?* and *Why can't we…?*

Role-play 3 was the last task in Week 4. I prepared two similar guided role-plays providing opportunities to use the two negative Q-forms. Each role-play provided four opportunities to use each Q-form, including one occasion for each negative Q-form, as in the dictogloss. In the first role-play, each student in pairs was given one of two task sheets, (A) or (B), on which a flowchart of a dialogue between a married couple (Mary and Tom) was printed, shown in Appendices 60 and 61, respectively.

The flow of conversation was very similar to one used in the Dictogloss. On Sheet (A) the instructions were given for Mary's lines and those for Tom were left blank, while on Sheet (B) the situation was reversed. In this guided role-play, the students playing Mary's role were supposed to ask Tom to do grocery shopping, while those playing Tom's role had to readily agree to the request. However, when Mary came home later that day, she noticed that Tom had still not done the shopping, providing Mary with an opportunity to produce a *Yes/no negative question* to express surprise, as in 'Haven't you done the shopping yet?'. Later that day, after doing the grocery shopping, Tom suggested to Mary that they buy a house next year, but Mary answered 'No, we can't'. This provided the students playing Tom to ask for reasons, using *Wh-negative question with 'why'*, as in 'Why can't we buy a house?'.

The other set of task sheets prepared for this role-play is shown in Appendices 62 and 63 in which similar, but not identical, instructions are provided. The students who played Mary in the previous role-play were given Sheet (A) and took Mary's role again (Appendix 62), but in this role-play they had an occasion to produce a *Wh-negative question with 'why'*. Those who were given Sheet (B) played Tom, in which they had the opportunity to produce a *Yes/no negative question*.

5.7.7.5 Questionnaire (W4)

The students filled out the same questionnaire used in the first three sessions (Appendix 25).

5.7.8 Summary of the four treatment sessions

As we have seen, the target rules were taught in the four treatment sessions using the different information-gap tasks. The number of target rules in each session was not the same, as summarised in Table 5.10 below.

TABLE 5.10 Summary of the four treatment sessions

<table>
<tr><th></th><th>Target Q-form</th><th>Dictogloss 1 (Listening & Modelling)</th><th>Dictogloss 2 (Listening & Reconstruction)</th><th colspan="2">Information-gap Task</th></tr>
<tr><td>Week 1</td><td>Q-tag</td><td>○</td><td>○</td><td colspan="2">Memory Game</td></tr>
<tr><td>Week 2</td><td>Cancel-inversion</td><td>○</td><td>○</td><td colspan="2">Guess Who</td></tr>
<tr><td rowspan="2">Week 3</td><td>Cancel-Inversion*</td><td>○</td><td>○</td><td colspan="2">Our neighbors</td></tr>
<tr><td>Q-tag*</td><td></td><td></td><td colspan="2">Do you have a good memory?</td></tr>
<tr><td rowspan="2">Week 4</td><td>Y/N-Negative</td><td>○</td><td rowspan="2"></td><td>Guided role-play 1</td><td rowspan="2">Guided role-play 3</td></tr>
<tr><td>Negative-Aux/Do-2nd</td><td>○</td><td>Guided role-play 2</td></tr>
</table>

Notes. In Week 3, the instruction focused on the use of Cancel-Inversion with the word 'there' as a subject of the subordinate clause and the use of Q-tag with the word 'there' as a subject of the sentence.

In the process of developing the teaching materials, I attempted to provide the students with comparable amounts of input and output for each target rule, but, needless to say, in the real classes it was not possible to control conditions as in experimental studies. For example, two stage 6 structures (*Cancel-Inversion* and *Q-tag)* were taught twice and the students had more opportunities to produce questions with these rules. Furthermore, although the students were given two dictogloss tasks in the first three sessions, in Week 4 they were not given the second dictogloss task including the reconstruction task. Moreover, the number of target questions in the recorded dialogues used in the dictogloss tasks also differed from session to session. These are points to be kept in mind during the data analysis of the effect of the treatment on individuals' acquisition of Q-forms, to be discussed in Chapters 6-8.

I should also note that, although the students were given the information-gap tasks in the first three sessions, in the last session the first two speaking tasks (i.e. Role-play 1 and Role-play 2) were more like exercises, in which they practised two negative Q-forms with the picture cards. As Ellis (2003) points out, designing tasks to provide opportunities to apply L2 features with a focus on meaning is not easy. In the tasks developed for this study, I attempted to provide occasions to produce questions with the higher-level rules, but I experienced difficulties with this. Some tasks in this study can be criticised as form-focused exercises (Role-play 1 and 2 given in Week 4), but from my observations of the students' performance while carrying out these tasks, I found that most students seemed to consider that these tasks were as useful as the others given in the treatment, since – through them – they got used to producing questions with the target rule (*Y/N-Negative* or *Negative-Aux/Do-2nd*), which many of them had not used before.

I will now show the materials developed for the three tests.

5.8 Materials development for testing

I gave one pretest and two posttests in the main study to examine the sustained effect of the instruction on individuals' advancement in developmental stage in interrogative acquisition. First, I will explain the rationale for the use of the tasks for testing. Secondly, I will show the tasks used in the tests and the procedures for their use. Thirdly, I will explain how I piloted these materials in order to make improvements.

5.8.1 Rationale for the use of the tasks for testing

I used two conversation tasks to collect examples of the students' questions. They were *Guided Role-play* and *Spot-the-Difference*, which will be shown later. There were three reasons for my choice. First, these conversation tasks had face validity as familiar classroom materials and could be designed to elicit sufficient question examples by seeding occasions to produce questions exemplifying the target rules.

Secondly, the tasks could be designed to promote guided yet relatively spontaneous production of various Q-forms including the target ones, which was necessary to assess learners' developmental stage and whether they had acquired each Q-form.

Thirdly, these tasks could be designed to be administered within the given time at each institution. I rejected the use of natural conversations, such as unstructured interviews, because the collection of enough examples of the target rules in such contexts seemed to require many hours and it was not feasible in the given time. Moreover, it was likely that many students would produce questions with less complex Q-forms in free conversations and the lack of examples of target questions was foreseen.

I prepared three sets of two kinds of similar, but not identical, tasks for the three tests. In each test the students were given the Guided Role-play first, and then the Spot-the-Difference task[30]. I will now show, in turn, these tasks and the procedures for their use.

5.8.2 The tasks developed for the three tests

5.8.2.1 Guided Role-play

The Guided Role-play was intended to examine the learners' ability to use questions from their conversation partner. The aim of these questions, 19 in number, was to collect information about specific accommodation in an English-speaking country, which they were supposed to share with either a host family or flatmates. The role of their conversation partner was played by an examiner who held the answers to the questions on the key sheet (see Appendices, 64, 65 and 66). The examiner's answers to the questions and other responses such as greetings and fillers were also strictly controlled to assure that every test-taker was given the same amount of information and the same quality of friendly responses.

In each test, on being provided with the instruction sheet (see Appendices 67, 68 and 69), the students were asked to read it through to see their role, and the context of the task, and how to carry it out. On the pretest (see Appendix 67), for example, the students had to imagine that they:

- ☐ were leaving Japan the following week to study English in a language institute in London;
- ☐ were renting a room in a flat which Mary, an elder sister of their American friend, had been sharing with some flatmates;
- ☐ had met Mary and stayed in her flat while visiting London with her younger sister the previous year;
- ☐ remembered what the flat was like, but needed to call Mary because they wanted to check if their memory was correct and ask further questions;
- ☐ had prepared and listed questions by themselves on a sheet, provided by the examiner.

In order to make the students feel at ease performing this role-play with the examiner, they were assured that Mary was willing to answer questions. They were also told that Mary was about five years older than them. This age was given because the examiner was actually older than the students and it was hoped that reminding them of the age difference would encourage production of polite forms of questions, such as 'Could/Would you tell me *if/Q-word* S+V?', in which one of the target rules (*Cancel-inversion*) was applied.

At the end of the instruction sheet, the following assessment criteria were provided in order to facilitate the students' appropriate and fluent use of English.

- ☐ The more you elicit the necessary information by asking questions appropriately and fluently, the higher you will score.
- ☐ Try not to fall silent. If you cannot ask questions in a full sentence, do your best still to ask questions. The examiner will answer your questions as long as they make sense, even if you produce questions in incomplete sentences.
- ☐ Try to use fillers and interjections where necessary to make a conversation lively and real.

The students read this instruction sheet, while waiting outside the testing room. Each student was invited into the testing room by the examiner and asked in Japanese whether they had any questions about the instructions.

Then, the examiner handed out a copy of the task sheet, on which were the instructions for a series of 19 guided questions (see Appendices 70, 71 and 72, for the task sheets used in the three tests). Using the task sheet used in the pretest (Appendix 70), I will illustrate how the students performed this role-play with an examiner. The examiner asked the students to read through the sheet to grasp the purpose of each question. The students were allowed to ask questions about the task sheet in Japanese, if necessary. On the sheet, some Japanese words were translated into English in order to help weaker students, so that they could concentrate on asking questions in order to elicit the necessary information from the examiner.

The examiner began the conversation by answering a phone call, following the instructions given on the answer key sheet (Appendix 64), as follows.

> Examiner: (Rrr.. rrr..) Hello?
> Student: Hello, Mary. It's [student's first name].
> Examiner: Hi, [student's first name]! How are you?
> Student: [Greet to Mary] Listen, I've got questions about the flat.
> Examiner: Sure, go ahead.

Each role-play was intended to provide the same number of occasions to apply each rule almost in the

same order, but the use of the target Q-forms was not obligatory and the choice of Q-form rested with the students. I will explain this in detail using the 19 task questions given in the pretest. The Q-forms I intended to elicit in the 19 task questions are shown in Table 5.11 below. In the table, the four target Q-forms are shaded.

TABLE 5.11 Q-forms intended to elicit in the Guided Role-play on the pretest

	Stage 3	Stage4		Stage 5		Stage 6		
	Fronting	*Y/N-Inversion*	*Wh-Inversion*	*Affirmative-Aux/Do-2nd*	*Negative-Aux/Do-2nd*	*Cancel-Inversion*	*Q-tag*	*Y/N-Negative*
Q1			○					
Q2	○							
Q3		○						
Q4				○				
Q5							○	
Q6							○	
Q7					○			
Q8							○	
Q9							○	
Q10			○					
Q11			○					
Q12				○				
Q13				○				
Q14			○					
Q15	○	○						
Q16					○			
Q17		○						
Q18					○			
Q19				○				
	2	3	4	4	3	?	4	0

We can see from the table above that two or more occasions to apply each rule were seeded in this role-play; except *Y/N-Negative* and *Cancel-Inversion*. For *Y/N-Negative*, although in the instruction this was one of the four rules I targeted, I eventually needed to give up on seeding occasions to elicit examples of this rule. This is because, for learners to be considered to have acquired this rule, at least two *Y/N-Negative* questions have to be produced with different auxiliary verbs in order to eliminate the possibility that they produced examples as a 'chunk'. In the role-play, however, it was difficult to seed enough occasions to elicit enough examples for this rule. For *Cancel-Inversion*, as explained earlier, the age difference between the students and Mary, whose role was played by the examiner, was set up to encourage production of polite forms of questions using this rule, but its use was optional.

It should be noted that, although I attempted to provide occasions to elicit examples for each Q-form, as shown in Table 5.11, it was possible to elicit the necessary information specified in the instructions given for each task question without using the target Q-form. In other words, the students' non-use of specified rules cannot be used as negative evidence for their non-acquisition of those rules. For example, the instruction for Q1 was intended to elicit a stage 4 *Wh-Inversion* question, as in 'How much is the rent?', but it was also possible to produce a *Cancel-Inversion* question, as in 'Could you tell me how much the rent is?'. Likewise, although the instruction for Q2 was intended to provide an opportunity to apply stage 3 *Fronting* in order to ask whether the rent included a gas bill, as in 'Does it include a gas bill?', the students could ask the same thing by producing a stage 4 *Y/N-Inversion*, as in 'Is a gas bill included in the rent?', or a *Cancel-Inversion* question, as in 'Do you know if the rent includes a gas bill?'. Furthermore, in producing questions for three task questions (Q10, Q11, Q14), the students could use either *Wh-Inversion* or *Affirmative-Aux/Do-2nd*. In Q11, for instance, both questions shown below are appropriate to ask the number of bedrooms in the flat.

How many bedrooms are there in the flat? (*Wh-Inversion*)
How many bedrooms do you have in the flat? (*Affirmative-Aux/Do-2nd*)

The use of *Cancel-Inversion* in this context was also appropriate Thus, in the *Guided Role-play* the choice of Q-form was ultimately up to the students, although with some guided questions the use of some Q-forms was inappropriate, and to some extent their options were limited.

In the examples given above, I included *Cancel-Inversion* questions, but I predicted that many students would not use this rule in the role-play, especially on the pretest, even if they had reached the highest stage. This is because the goal of this role-play was to elicit the specified information from their conversation partner and it is natural for learners to incline towards easier, familiar features in producing questions under communicative pressure. Furthermore, although Mary's age was set four or five years older than the students, the created context might not be sufficiently formal for them to use this rule, since they knew her and had stayed with her before[31].

While developing this task, I also realised the difficulty of eliciting examples of *Q-tag*. In each role-play, I attempted to seed four occasions to encourage the students to apply this rule. For example,

in the pretest Q5, Q6, Q8 and Q9 were of this type. On the task sheet given to the students, they read the instruction, which was shown above the instruction for Q5, as follows:

> You remember the following four things (marked with ①-④), but your memory is not clear. Ask Mary a question for each in order to check if your memory is correct.

By giving this instruction, in Q5 I intended to provide occasions to produce a question with a *Q-tag* , such as 'The flat is near the language centre, isn't it?'. However, I predicted that many students would not do so in the given context for two reasons. First, the students were not asked to use *Q-tag* in the instructions; in other words, its use was not obligatory. Secondly, many students did not seem to be aware that the use of tag questions was available for checking whether something is true, or for asking for confirmation.

For *Negative-Aux/Do-2nd*, three occasions to apply this rule were seeded in each role-play. In the pretest Q7, Q16, Q18 came under this category. For example, in Q15 the students were required to ask if there was a washing machine in the flat and heard the reply that 'Yes, there is, but we can't use it at the moment', and then they were instructed in Q16 to ask a reason for inconvenience, which was intended to produce an occasion to use *Negative-Aux/Do-2nd*, as in 'Why can't we use it?'. Many students who volunteered in my piloting produced such a question on the three seeded occasions, but it was also possible to ask simply 'Why?', as native English speakers may do, giving priority to fluent conversation. So the use of *Negative-Aux/Do-2nd* in these three occasions was also not obligatory.

As discussed in Chapter 3, for a learner to be considered to have acquired rule *x*, at least two productive examples of rule *x* have to be present in that learner's sample. I expected that many students would try to use the target rules on the posttest and follow-up test, but it was likely that this would not be the case in the pretest. Non-production of rule *x* in the task does not necessarily mean that learners have not acquired this rule, unless the use of rule *x* is obligatory. To make sure that I could collect enough examples of target rules from each student, I decided to seed obligatory occasions in the Spot-the-Difference task. I will explain the procedures for this in the next section.

5.8.2.2 Spot-the-Difference task

The Spot-the-Difference task consisted of three parts:

(a) five-pieces of-missing-information (Part 1),
(b) eleven-differences, (Part 2) and
(c) confirm-the-differences (Part 3).

Each part was intended to provide obligatory occasions to use specific Q-form(s) in order to obtain the necessary information from the examiner. It should be noted that no occasions to elicit examples of stage 6 *Y/N-Negative* were seeded in this task, for the same reasons, as in the Guided Role-play.

I prepared three similar, but not identical tasks, for the three tests. I will now illustrate how examples of the target rules were elicited in each part by using the materials used in the pretest.

5.8.2.2.1 Part 1: Five-pieces-of-missing-information

Part 1 (Five-pieces-of-missing-information) was intended to provide two obligatory occasions to use *Negative-Aux/Do-2nd*. In this part, the students were asked on their task sheet (Appendix 73) to elicit five pieces of missing information in the five speech bubbles, which were numbered from ① to ⑤. The examiner answered questions from the answer key shown on the task sheet (Appendix 74). Three of the five task questions were distractors (①②④), although they provided the students with occasions to apply non-target rules, such as *Wh-Inversion* or *Affirmative-Aux/Do-2nd*. In the other two task questions (③⑤) the students were required to ask reasons for something. In the task question marked with ③, for instance, they saw an incomplete sentence 'I can't eat dinner at home because ③ ' in the speech bubble and had to ask a reason in order to fill the blank, which provided an occasion to produce a *Negative-Aux/Do-2nd* question like 'Why can't the boy eat dinner at home?'. For the task question marked with ⑤, they saw an incomplete sentence 'I haven't cleaned the bathroom because ⑤ ' in the speech bubble, and they were expected to apply this rule, as in 'Why hasn't the man cleaned the bathroom?'.

Before starting Part 1, to eliminate the production of one-word questions to fill these two blanks, as in 'The boy can't eat dinner at home. Why?', the examiner instructed the students to ask questions in sentence form, as cited below.

> Examiner:
> Now, you are going to ask five questions to elicit five pieces of missing information in the speech bubbles shown on your task sheet. They are numbered from ① to ⑤. Ask five questions in order, starting with ①. You should ask a question for each *in a sentence*. Only when you can't do so, may your questions be in a phrase or a single word.

Again, the application of *Cancel-inversion* was possible to elicit the missing information for these two blanks (e.g. '③Could you tell me why the boy can't eat dinner at home?'), but if the students applied this rule in producing questions to fill these two blanks, they were instructed by the examiner to ask direct questions rather than indirect questions, as in 'Don't start your questions with *could you tell me*'.

Thus, the students were not allowed to use *Cancel-Inversion* here, but it should be noted that it was possible for the students to apply a different rule in asking questions to fill these two blanks. For example, some might apply *Wh-Inversion* to elicit the missing information in blank ③, as in 'What's the reason for the boy's eating out tonight?', although this question did not sound natural. In other words, the application of *Negative-Aux/Do-2*nd for these task questions was not *obligatory*. However, it was expected that there was a good chance in Part 1 of obtaining either positive evidence of *Negative-Aux/Do-2*nd or negative evidence, such as 'Why the boy can't eat dinner at home?' in which the auxiliary 'can't' was mis-placed.

Before starting this task in each test, in order to help the students understand how to ask questions in Part 1, a sample question was provided on the instruction sheet (Appendix 75). It should be noted that similar information was missing in the five speech bubbles shown on the task sheet used across the three tests. The incomplete sentences shown in the five speech bubbles in each test are summarised in Appendix 76. The task sheet given to the students in the two posttests are shown in Appendices 77 and 78 and those by the examiner are in Appendices 79 and 80.

5.8.2.2.2 Part 2: Eleven-Differences

Immediately after the production of five questions in Part 1, the examiner moved to Part 2 (Eleven Differences). This part comprised two components:

(a) eliciting eight questions without any instructions and
(b) eliciting examples of *Cancel-Inversion* with the instructions.

First, the students were told by the examiner that there were 11 differences between the picture on their sheet and the one on the examiner's task sheet. They were instructed to ask questions in a sentence to spot as many differences as they could within five minutes. In both pictures the same four people and one pet were drawn with many other objects, but some of their actions and the number/colour of some objects were different between the two pictures; these differences constituted the information gap. For example, in the student's picture used in the pretest (Appendix 73), the woman in the kitchen is wearing a pair of red earrings, while in the examiner's picture the earrings are white. The differences seeded between the two pictures in the pretest are summarised in Table 5.12 below.

TABLE 5.12 Eleven differences in Part 2 of the Spot-the-Difference task (pretest)

		Student's picture	**Examiner's picture**
Woman in the kitchen	(1)	Wearing a pair of red earrings	Wearing a pair of white earrings
	(2)	Holding a spoon in her hand	Holding a pair of chopsticks in her hand
Boy on the floor	(3)	Not yawning	Yawning
Woman at table	(4)	Wearing a hat	Not wearing a hat
	(5)	Wearing a blue T-shirt	Wearing a white T-shirt
Man at table	(6)	Drinking a glass of beer	Drinking a glass of water
Cat	(7)	Sleeping on the table	Sleeping on the floor
Clock	(8)	Showing nine o'clock	Showing eleven o'clock
Ball under the table	(9)	Soccer ball	Volley ball
Trees outside the house	(10)	Two trees	Three trees
Newspaper	(11)	*Weekly Times*	*Daily Times*

In Part 2, I provided the students with eight occasions to ask questions without any instructions to apply specific rules in order to spot the differences. They indicated the differences they had spotted on their task sheet. Although few students would voluntarily use the target rules, and I might not be able to collect many target questions, especially on the pretest, it was hoped that those who had taken the instruction might give it a try and produce target questions on the posttests.

After eight questions, the students were told by the examiner to stop working, as in 'Okay. Time's up! Put down the pen and wait for a moment'. The examiner then checked the number of *Cancel-Inversion* questions produced by a student in Part 1 and 2. If there were two or more instances, no instructions to elicit further examples of this rule were provided. If there was only one or no example, the examiner gave the instructions to provide obligatory occasions to apply this rule, which will be explained in detail in the next section. This evaluation work was carried out with notes taken on the examiner's evaluation sheet (Appendix 81). While administering this task, the examiner was required to check whether or not a Q-form was appropriately applied in each example, where I interpreted *appropriateness* in terms of *word order*. If a student applied a Q-form appropriately, the examiner put a circle '○' in boxes provided on the right of the evaluation sheet for each rule. If a student failed to apply a Q-form, the examiners put a cross '×'. On the sheet three target Q-forms and four non-targeted rules at stage 3 and above were listed with sample questions. *Y/N-Negative* was not included on this list because no occasions to apply this rule were seeded in this task and it seemed unlikely that the student would produce questions with this rule.

I should point out that this evaluation was not always straightforward. For instance, question utterances produced by weaker students tended to be muddled with many pauses, as in 'Woman kitchen mm . do woman earring… doing kitchen', and it was hard to identify a Q-form intended by the speaker. Moreover, with incorrect question utterances starting with a *wh-* or *Q*-word, such as 'What boy is colour pants?', the examiner could construe more than one question with a different Q-form, such as *Affirmative-Aux/Do-2nd* questions (e.g. 'What colour of pants is the boy wearing?') and *Wh-Inversion* questions (e.g. 'What is the colour of boy's pants?'), and not be able to identify which rule the speaker meant to use.

In order to reduce the examiners' load in giving this task, I told them that they did not need to worry about precision in identifying rules at stage 5 or below, since many occasions to apply each rule at stage 5 and lower were seeded in the first task (*Guided Role-play*) and we would not lack both positive and negative evidence for these rules.

5.8.2.2.3 Elicitation of *Cancel-Inversion* questions in Part 2

I will now explain how the examiner instructed the students to produce questions with *Cancel-Inversion* after eight questions in Part 2. In the elicitation, the examiner first gave a student the following instruction in Japanese

> Let me give you a little bit more time to find more differences. Please ask questions starting with 'Could you tell me…?'.

From the second elicitation, the examiner instructed the students, as follows, until two pieces of either positive or negative evidence were produced.

> Okay. Find one more difference by starting a question with 'Could you tell me…?' again.

We should note that if a student produced *Y/N-Inversion* questions starting with the given clause, as in 'Could you tell me the price of the hat?', the examiner did not put any mark on the evaluation sheet and instructed that student to find a difference in terms of one of the characters' actions by starting a question with 'Could you tell me', as follows:

> Okay. Then, find a difference in one of the characters' action by starting a question with the same clause, 'Could you tell me'.

This instruction was intended to facilitate the production of indirect questions with *Cancel-Inversion* like 'Could you tell me what the boy is doing in your picture?'. These instructions were printed on a copy of the evaluation sheet (Appendix 81). It should be noted that the examiner did not use the elicitation method if a student had produced a total of two questions in which he/she had failed to apply

Cancel-Inversion in Part 1 and 2, since these instances could be used as negative evidence for their non-acquisition of this rule.

A similar elicitation method was used in Spada and Lightbown (1993) to elicit *Wh-Inversion* questions targeted in their instruction, as mentioned in Section 3.5.6. In this study learners were asked to find more differences by starting their question with 'what' when they did not spontaneously produce a wh-question after two questions, as in 'Can you ask me a question with *what*?' (ibid.:211).

After eliciting examples of *Cancel-Inversion* at the end of Part 2, the examiner also checked if a student had produced two questions for stage 4 *Wh-Inversion* and stage 5 *Affirmative-Aux/Do-2nd* in Parts 1 and Part 2. If a student had not produced two *Wh-Inversion*, the examiner instructed that student to find more differences by starting their questions with either 'where' or 'how'. If a student had not produced *Affirmative-Aux/Do-2nd* questions, the examiner instructed her/him to find more differences in the characters' actions by asking questions with 'what'. As mentioned earlier, although both the *Guided Role-play* and the *Spot-the-Difference* were seeded with many occasions to apply these rules, the choice of Q-forms was up to the students. So I predicted a lack of examples of these rules among weaker students, which might cause problems in the assessment of their developmental stage.

5.8.2.2.4 Part 3: Confirm-the-Difference

After eliciting questions in Part 2, the examiner asked the students 'How many differences have you found?'. Then the examiner took notes of their reply on the evaluation sheet and confirmed it, as in 'Five differences. Well done!'[32]. This question was intended: (a) to remind the students that the task goal was to find as many of the eleven differences as they could and (b) to pave the way to carrying out the elicitation of examples of *Q-tag* in the next part (Part 3). The examiner then prompted that student to produce an example of a *Q-tag* as follows. This was intended to provide the students with three obligatory occasions to apply this rule.

> Before ending this task, let me give you opportunities to check whether some of the differences you've found are correct. Now, choose one difference in which you're not sure if your answer is correct and ask me a question. If you find a mistake in your answer, you can make changes.

> Let me give you an example. If you think that the price of the tennis racket in my picture is 170 dollars, but are not sure, you may ask 'In your picture, the price of the tennis racket is 170 dollars, isn't it? (↑)'. Start your question with 'In your picture'.

By giving an example in which a *Q-tag* was used, I intended to make the use of this rule obligatory. If a student did not apply this rule, the examiner repeated the sample question with an emphasis on the tag to remind that student that he/she was supposed to ask a question by using a *Q-tag*. Again, the use of such instructions might draw the students' attention to *form*, but in Part 3 they were supposed to produce questions to check whether or not their answers were correct; and it is possible to say that their attention was also drawn to *meaning*.

After eliciting one example, the examiner instructed that student to check whether another difference they had found was correct, as in the following:

> Well, you can ask one more question to check whether another difference you've found is correct. Start your question with 'In your picture' again.

Finally, the examiner further instructed that student to repeat the exercise, but this time the student was asked to choose one difference for which he was sure that his answer was correct, just to satisfy himself, as follows:

> Now, I want to give you one more opportunity to confirm whether another difference you've found is correct. This time, however, choose one difference in which you think that your answer is correct. Start your question with 'In your picture' again.

These instructions were intended to provide the students not only with three occasions to apply a *Q-tag* but also with opportunities to produce tag questions with the appropriate intonation. As mentioned earlier, I taught the use of intonation in tag questions in Week 1 and, although I did not plan to analyse the students' ability to use the correct intonation in this study, I included these two different instructions, since I thought that would remind them in tag questions of the use of intonation. It should be noted that the examiner provided all students with three occasions to use a *Q-tag*, even if they had voluntarily produced two tag questions in Part 1 and Part 2. In my study, the tests were intended not only to collect data for the analysis but also to give the students opportunities for using what they had been taught, and

the provision of three occasions to produce tag questions at different levels of certainty was intended to remind them of the correct use of intonation, which had been focused on in the mini-lessons in Week 1 and Week 3.

5.8.3 The examiners

In order to administer each test within the given time at each institution, the Guided Role-play and the Spot-the-Difference were given by four examiners (Examiner A, B, C, D), including myself (Examiner D). All examiners were female, teaching English part-time at different institutes at the time. I will briefly introduce each examiner's background.

Examiner A was doing a part-time TESOL master's course at a university while teaching English to a group of Japanese businessmen at a language school. She was in her early 50s and had taught English for about four years. Examiners B and C had taught English at several secondary schools for about five years on a part-time basis, and were in their 30s and 40s, respectively. These two examiners used to be my colleagues when we were teaching English at the same senior secondary school for about a year. Three of us had a certificate for teaching English at secondary schools. Examiner A and B had lived in the USA for a year with their family, while Examiner C had not. Examiner D was in her 30s then and had taught English at a senior secondary school and other institutes for about five years in total, including the nursing college where I carried out my main study. When I left Japan to study at the University of Edinburgh in 1999, Examiner C replaced me at the nursing college, although she had not taught the students in this study before the main study. As mentioned earlier, I had an assistant teacher in the four treatment sessions at the nursing college; Examiner C taught Group A in the first three sessions and Examiner B in the last session. So only Examiner A was not involved in the treatment sessions at the nursing college. At the university, I taught all four sessions without any assistant. So Examiner A, B and C shared a lack of familiarity with the university students. It should be noted that Examiner C had administered the preliminary tests on my behalf at the two institutions, which suggests that the students at the two institutions had more contact with her than with Examiners A and B.

Originally, these four examiners were scheduled to give the two tasks in the three tests at both

institutions, but Examiner C became ill a few days after giving the follow-up test at the university and could not join us at the nursing college to give the same test. I then asked another friend (Examiner E) to replace Examiner C. Examiner E was in her late 20s and had a certificate for teaching English at secondary schools. She had taught English at private English schools on a part-time basis for about one year. Her teaching experience was shorter than that of the others, but she had spent three years in the USA and attended a local high school, while staying with a local family. Thus, the five examiners had different backgrounds in terms of their age and experience in teaching and living abroad.

5.8.4 Training sessions for the examiners

I gave the three examiners (A. B, C) two training sessions before the pretest and one session before each posttest. Examiner E took one session separately before the follow-up test. The purpose of the sessions was to familiarise the examiners with the testing materials and the procedures for giving the Guided Role-play and the Spot-the-Difference. I will show the content of each session below.

First training session

In the first training session, I explained the purposes of my study and the 6-stage sequence in English interrogative acquisition and told the three examiners that they had to give two conversation tasks in each test. I then provided a copy of the instruction sheet for the Guided Role-play used in the pretest and asked them to read it through. I then gave each examiner a copy of the task sheet given to the students in this task in the pretest (Appendix 70) and asked them to read through the instructions for the 19 task questions shown. I told the examiners that the purpose of this task was to collect 19 questions with different forms from each student and explained which rule was intended to be elicited in each task question. Then I played a videotape, on which they saw me administering this task to a Japanese volunteer in Edinburgh. During the first viewing, I encouraged the examiners to get an idea of the task. Provided with a copy of the answer key sheet (Appendix 64), the examiners were asked to go through the 19 questions and answers while listening to the recorded conversation without watching the video. I reminded the examiners that, in this task, the choice of Q-forms in the questions was ultimately up to the students and that the examiners should not encourage the students to use specific Q-forms. I also

reminded them of their two different roles in the task as examiner as well as conversation partner who was an elder sister of the student's friend, and encouraged them to be friendly to the students. Then, four of us performed the role-play twice in pairs, changing roles.

I moved on to the Spot-the-Difference task and gave the three examiners a copy of the instruction sheet used in the pretest. After reading it through, they were provided with a copy of the task sheet which was given to the students in the pretest (Appendix 73) and studied the picture drawn on their sheet. I then performed the task, taking the role of examiner, with one of the examiners taking the student role. The other two observed our performance while looking at their task sheet from time to time. After giving Part 1 (Five-missing information), I explained that the main purpose of this part was to elicit two stage 5 *Negative-Aux/Do-2nd* questions on two occasions (③⑤), and that if a student produced a *Cancel-Inversion* question on these two occasions, the examiner had to instruct that student to ask a question without starting it with 'Could you tell me…?'.

In Part 2 the examiner performing the student's role asked eight questions and took notes of differences she had found on her task sheet. The other examiners also took notes on their own task sheets of the differences which that examiner seemed to have found. No *Cancel-Inversion* questions were produced by the examiner performing the student's role in Part 1 and 2, so I gave her the instructions intended to elicit this rule, and she produced two examples correctly. I then gave each examiner a copy of the evaluation sheet (Appendix 81) and explained how to carry out the identification of Q-forms in the students' question utterances, and how to put the marks in boxes printed on the evaluation sheet. I showed them my copy of the evaluation sheet which I was using, on which they saw 17 circles in the boxes provided. The examiner performing the student's role had appropriately applied rules at stage 4 and stage 5 in producing five questions in Part 1, eight questions in Part 2, and *Cancel-Inversion* in producing two examples in the elicitation. So I did not need to elicit stage 4 and stage 5 questions by giving the instructions, but I explained when and how to carry out the elicitation of examples of stage 4 *Wh-Inversion* and stage 5 *Affirmative-Aux/Do-2nd*.

I then asked the examiner performing the student's role how many differences she had found so far and took notes of the number reported on the evaluation sheet and moved to Part 3 (Confirm-the-difference).

I elicited three examples of *Q-tag* by giving the instructions printed on the evaluation sheet.

To familiarise the three examiners with the elicitation procedures, I performed the student's role and asked another examiner to perform the same task as an examiner. The other two examiners did the evaluation work while observing our performance. I made mistakes on purpose in all three parts, so that the examiner had to give me all the instructions which I had explained to them earlier in the training session. After the task, the three examiners told me that the evaluation work was difficult because they were sometimes not sure which rule was used in questions. I then told them that they did not need worry too much about precision, since many occasions to use different rules at stage 5 and lower were seeded in Part 1 and Part 2. But I emphasised the importance of giving the prescribed instructions to provide the necessary number of occasions to apply *Cancel-Inversion* and, where necessary, other non-target rules (i.e. *Wh-Inversion*, *Affirmative/Aux/Do-2*nd), at the end of Part 2 and three occasions to use *Q-tag* in Part 3. Four of us then performed the same task in pairs twice, changing roles.

The first training session took about two and half hours. At the end of the session, I provided them with extra task and evaluation sheets, so that they could practise the examiner's role before the next training session scheduled for about one week before the pretest.

Second training session

After the first session, I asked a native English speaker to perform the two tasks used in the pretest with me, so that I could audiorecord our conversations. I played the student's role and the native speaker played the examiner's. I made this recording because, in the first training session, I had observed that the way each examiner had responded to the students' questions from the same key sheet differed in some aspects of language use, such as accent, speed, and tone of voice. In this session I played this recording twice and asked the examiners to learn how they had to answer questions from the students. I then gave the examiners a copy of this recording.

The examiners asked me a few questions about how to do the evaluation work and I answered them. We performed the Spot-the-Difference task twice with the same task sheets, changing roles. I then

explained the seating arrangements for the pretest, which will be shown in the next section, and showed them how to use the two audiorecorders assigned to them in the pretest. The second training session took about one and a half hours.

Third Training session

The third training session was given about one week before the posttest. By using the materials prepared for the posttest, four of us performed each task twice for each, changing roles. Since giving the pretest, the examiners had become more adept at giving the tasks. In the pretest I noticed a problem, in that many students at the nursing college took a long time to produce eight questions in Part 2, the Spot-the-Difference. So I instructed the examiners to move to the elicitation of examples of *Cancel-Inversion* and the other rules after four questions if a student spent about five minutes on Part 2. The third training session took about one hour.

Fourth training session

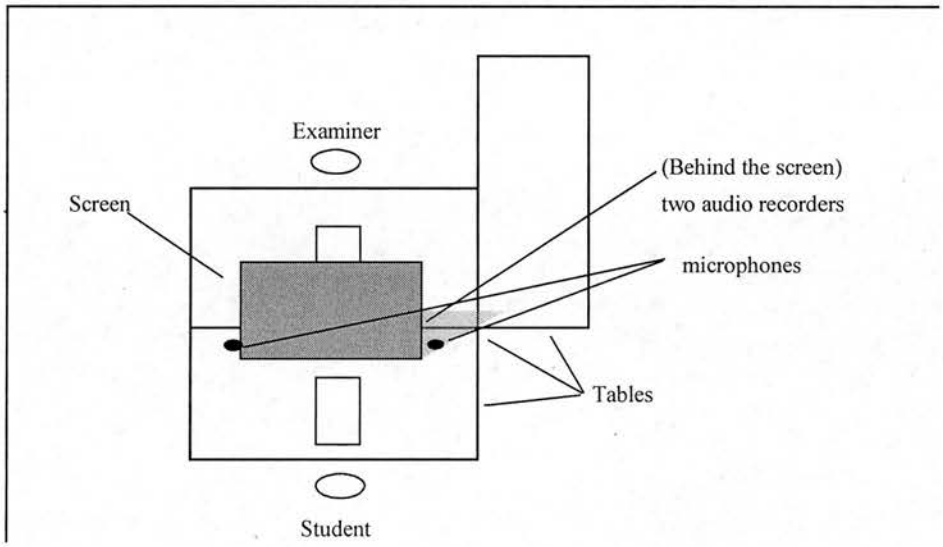
The three examiners (A, B, C) were given the fourth session about one week before the follow-up test and practised the two tasks with the relevant materials. I gave Examiner E, who replaced Examiner C, a one-to-one training session once, in which I explained the purpose of my study, the 6-stage sequence and the procedures for each task. We performed both tasks twice for each, changing roles. Examiner E gave this test only to the nursing college students, who tended to speak English slowly, and she did not have difficulties in carrying out the evaluation work while answering questions from the students.

In addition to these training sessions, before giving each test at each institution, I had a short meeting with the examiners to remind them of the important points in giving the tasks, such as providing obligatory occasions to apply specific rules in the Spot-the-Difference task.

5.8.5 Seating arrangements

At each institution, the four examiners administered the two tasks in separate rooms. The size of rooms varied approximately from 25 to 50 square meters. In order to provide the students with similar conditions, desks, chairs, and recording equipment were placed in the same position in each room, as shown in Figure 5.2.

FIGURE 5.2 Seating arrangements for the testing rooms



In each testing room, a student and the examiner sat face to face across the desks. These were placed in the furthest corner of each testing room, which was farthest from the entrance, and the students sat facing the wall, so that they could concentrate in a quiet environment. During the two tasks, a shoulder-height screen was put between the two, but there was eye contact. The purpose of the screen in the Guided Role-play was to create distance between the two, who were supposed to be talking on the phone, but I encouraged the examiner to speak to the students in a friendly way, with some eye contact, to create a relaxed atmosphere in the test rooms. In the Spot-the-Difference, the screen was used to hide the examiner's task sheet and evaluation work while answering questions from the students. Two audio recorders were placed behind the screen on the examiner's side, so that they were invisible to the students, who could only see two microphones[33].

5.8.6 Questionnaires given after the three tests

Immediately after completing the three tasks in each test, the students were asked to fill out a questionnaire. Its purpose was: (a) to give the students an opportunity to reflect on their test performance and (b) to collect their self-reports on how their production of target questions had changed in each test and outside the classroom. The questions included in the questionnaire given after the pretest were also included in the other two questionnaires given after the post- and follow-up tests.

5.8.6.1 Common questions included in the three questionnaires

In the questionnaire given after the pretest (Appendix 82), the students were asked:

- ☐ to indicate their perceptions of task difficulty for each task on a five-point scale;
- ☐ to name the most difficult task among the three and to give reasons;
- ☐ to name the easiest task among the three and to give reasons;
- ☐ to report mistakes they had noticed while carrying out the tasks in the tests; and
- ☐ to report whether they think they managed to make corrections when they noticed those mistakes.

In addition, at the end of the each questionnaire, a space was provided to encourage the students to write comments and suggestion, which they had not had a chance to express in the other part of the questionnaire.

5.8.6.2 The questionnaire given after the posttest

In the questionnaire given after the posttest (Appendix 83), in addition to the above questions, the students were asked to report:

- ☐ any positive changes they had noticed in their use of the expressions to facilitate conversations in pair work taught in the treatment, and
- ☐ any positive changes in their noticing mistakes in producing questions on the posttest.

5.8.6.3 The questionnaire given after the follow-up test

The follow-up test was given either six or seven weeks after the last treatment session. I was interested in examining if the instruction had a long-term effect on the students' production of target questions in and out of L2 classes and on other aspects of L2 learning. So in the questionnaire given after the follow-up test (Appendix 84), I asked the students to report:

- ☐ any conscious efforts which they had made in producing questions in each task,
- ☐ any positive changes in producing questions in each task, and
- ☐ any positive changes in using the target rules in their daily life after the treatment.

5.9 Piloting earlier versions of materials for the main study

I piloted earlier versions of the teaching and testing materials with 11 Japanese volunteers who were taking the *General English* course at the Institute for Applied Language Studies (IALS) at the University of Edinburgh in August and September in 2001. More than two thirds of the volunteers were preparing for studying at a British university from October 2001. The English proficiency of some of them seemed higher than that of most of the students in my main study. Most of the volunteers were leaving Edinburgh soon and I had to complete the piloting within a month. In the next section, I will explain the purpose of piloting earlier versions of testing materials and the procedures.

5.9.1 Piloting earlier versions of testing materials

There were three purposes for piloting the earlier versions of testing materials. They were to see:

(a) if the instructions for the three tests and for the four sessions of teaching were clear and easy to understand;

(b) if the procedures for eliciting question examples of the target rules in each task in the three tests and in the four treatment sessions were effective; and

(c) if the evaluation procedures for the Spot-the-Difference task were clear and easy for the examiners

to use.

I wanted to give the 11 volunteers three tests as in the main study, but many of them were pressed for time because they were attending a daytime course. So I gave each volunteer only one of the three tests before piloting the teaching materials with these students. I made an appointment to meet up with each volunteer when they were free between the 20th and 24th August. While giving the volunteers the three tasks, I recorded our conversations. However, I did not transcribe the recordings for detailed analysis, with the main study in Japan close at hand, and the need for time to analyse the two tests given before the main study and to catch up with the delay in my preparation, which was caused by the unexpected problem mentioned earlier.

While giving the tasks to each volunteer, I took notes on my task sheet when I noticed problems with the procedure and in the materials. I will now summarise my observations of the volunteers' use of target rules in the two conversation tasks.

In both tasks, most volunteers did not voluntarily produce stage 6 questions, but two produced *Cancel-Inversion* questions without any instructions. After the test, one of them told me that she consciously produced indirect questions starting with 'could you tell me' in speaking, since she remembered that she had been told by her English teacher back in Japan that the use of such indirect questions would convey the speaker's politeness and that the use of such indirect questions in speaking tests like TOEFL would be regarded as evidence of a high level of communicative competence. She also said that, since then, she had started making an effort to use this Q-form in speaking and had become able to use it. Her comment reminded me of the importance both of raising learners' awareness of the usefulness of L2 structures in communication through instruction, and of providing opportunities to use them, which could help learners increase linguistic options.

In the Spot-the-Difference, all volunteers tried to apply stage 5 *Negative-Aux/Do-2*nd questions in the two seeded occasions to apply this rule in Part 1. Most volunteers also tried to apply *Cancel-inversion* in the seeded obligatory occasions to apply this rule. In Part 3, in which the examiner gave the instructions to ask questions using *Q-tags*, most volunteers tried to use this rule in the given three

occasions. Some volunteers used other Q-forms in the first elicitation, but noticed that they were supposed to produce tag questions after hearing the further instruction. This confirmed for me that the instructions for the elicitation of these rules were effective and indispensable for collecting examples of the target rules. These observations clearly show that the absence of examples of these target rules in learners' output does not necessarily mean that they have not acquired them. This highlights the need to carry out developmental stage assessment with concrete data.

In the monologue task (Picture-cued Production), all volunteers completed the task within five to ten minutes and did not experience any difficulty in understanding the pictures either shown on the cards or the task procedures. After this task, two students mentioned that they had not used tag questions for communication, but learnt through doing the task which required them to produce tag questions. One of them also said that it was the first time she had produced *Y/N-Negative* questions like 'Haven't you done the ironing yet?', although she knew how to form them, and that she learnt from the picture given on the card that the pictured person was expressing his/her surprise by producing a *Y/N-Negative* question.

After the piloting, I revised the materials by using notes I took while giving the tests and comments and suggestions given by the volunteers. For example, I improved the examiner's evaluation sheet used in the Spot-the-Difference; I had difficulty in giving the instructions to apply specific Q-forms smoothly which were printed not on the evaluation sheet but on a separated sheet, and integrated the instructions into the evaluations sheet.

5.9.2 Piloting earlier versions of teaching materials

After piloting the testing materials, I piloted earlier versions of teaching materials for the four sessions. The materials for each session were designed to be taught in an 80-minute class and it would have been ideal to spend the same length of time on piloting them, but there were only two dates available to get together. On each date, two and a half hours was available and I piloted materials for Week 1 and Week 2 on the first available date, and those for Week 3 and Week 4 on the next available date. In order to pilot most materials within the given time, I showed the peer-assistance version of the modelling

video prepared for Week 1 and Week 3 only, and shortened each speaking task.

At the beginning of the first meeting, I told the volunteers that the two lessons which they were to be given that day were designed for the students in my main study, and that my main focus was to pilot them in order to improve the materials and teaching procedures. I also explained that, due to limited time, I would cut short my teaching, so that I could pilot most materials prepared for two 80-minute lessons in two 60-minute classes. I encouraged the volunteers to give me comments on any aspects of the lessons after each class. While giving the instructions, I jotted down notes on a sheet of paper when I noticed problems.

None of the 11 students complained about the lack of immediate teacher feedback on their mistakes while performing the speaking tasks. For the requirement to pay attention to output, many said that it was difficult to do so when they were performing the information-gap tasks, although it was easier in the dictogloss reconstruction task. They also said that they were pleased to receive peer feedback, and that the modelling video made them feel more confident to give their partner cues and hints when they made mistakes. Furthermore, teaching the volunteers enabled me to understand how much time learners would take to complete each task. For example, in some instances I spent more time than I had planned and notes taken during and after the instruction helped me improve teaching procedures.

For the modelling video, many volunteers told me that they had never seen this type of modelling (i.e. performed by a learner pair) and were motivated to try out some phrases used in the modelling when helping their partner to notice mistakes and make self-corrections. Many students also said that they had not received instruction which required them to use higher-level Q-forms. Some volunteers also said that they knew the target rules, but had not consciously tried to produce questions using them. A few students went on to say that, since they had arrived in Edinburgh to study English, they had heard native speakers producing tag questions and wanted to be able to reproduce them like native speakers, but hesitated because they had lacked confidence. One student admitted that he had not felt an immediate need to use the target rules in Japan, but now recognised their usefulness.

These positive comments gave me confidence to carry out the main study in Japan and I confirmed the

need to provide opportunities to practise the target Q-forms in communicative tasks with a focus on form. Apart from positive comments on the instruction, some volunteers pointed out several handouts which required revision and I integrated their suggestions into the final version of teaching materials.

5.9.3 The final research design

After finishing piloting and revising most of the materials for testing and teaching, I heard from the lecturer at the university that I could carry out my main study with her two classes as I had planned, but I foresaw a possibility of a decrease in the number of students. I considered eliminating the two kinds of additional data collection (i.e. recording classroom conversations and giving the monologue task) from my main study, but eventually decided to retain both of them, since their implementation was feasible. However, I decided not to analyse the students' interaction in the classroom in this study because it would involve much transcribing and was beyond the scope of the study. For the same reason, I eventually decided not to analyse question examples collected in the monologue task in this study.

5.10 Summary

I have explained in detail the content of the noticing-promotion instruction and the procedures for each component of the instruction, so that other teachers can duplicate the tests and instruction where necessary. In the next three chapters, I will present an analysis of the results. In Chapter 6, I will show the results of the examination of the effect of the noticing-promotion approach on individuals' change in developmental stage in the acquisition of question formation, which was this study's first aim. In Chapter 7, I will present the result of my examination of the effect of the instruction on filling gaps in individuals' acquisition of the target rules in relation to their readiness to learn them, which was my second aim. In Chapter 8, I will show the results of my examination of the effect of the instruction on the students perceptions of the usefulness of the instruction and of their changes in producing questions with the target rule, and on their classroom behaviour in pair work, which was my third aim.

CHAPTER 6
Effects of the treatment on individuals' gain in developmental stage

The first aim of the study is to examine the effects of the treatment on individuals' gain in developmental stage in the acquisition of interrogatives. To this end, I formulated the following research questions:

Q1. Is the *noticing –promotion approach* effective in helping the students advance in developmental stage in interrogative acquisition?

Q2. If so, are there any differences in individuals' stage gain in relation to their current developmental stage?

In Section 6.1 I will explain the transcribing method of the recordings of the students' performance in the three tests. In Section 6.2 I will explain the coding system of question examples. In Section 6.3 I will explain the procedures for pretest stage assignment. In Section 6.4 I will explain the procedures for the measurement of individuals' sustained gain in developmental stage on the two posttests. In Section 6.4 I will present the results of my analysis with discussion.

6.1 Transcribing method

I obtained a total of about 90 hours of recordings of the students' question utterances, collected in the three tasks given on the three tests. Transcribing took about six months in total from May until November 2002[34]. I transcribed the recordings orthographically unless the students' pronunciation differed markedly from standard English. Where there were marked differences, such as the use of Japanese words and fillers, I transcribed utterances by using the Hepburn system of Roman letters in *italics*, as shown in Table 6.1 on the next page, followed by an English translation typed also in *italics* in square brackets. Where English words were mispronounced and differences marked, I delimited them with slashes. Periods of silence were indicated by dots: one dot denotes a period of silence lasting less than one second; three dots denote one lasting more than one second. Where students' utterances were inaudible or indecipherable, I indicated that as '[inaudible]'.

I did not transcribe the examiner's utterances because the examiner replied to students' questions by using the answer sheet provided for each task. Where the examiner's utterances went beyond the set answers or where their reply overlapped with students' utterances, such as when repeating what a student had said in order to encourage them to keep talking, I transcribed these and enclosed them in round brackets. Where a pair of round brackets appears between students' two utterances without any transcripts (e.g. 'Why? ( ) Did you forget?'), it indicates that the student continued a question relevant to the previous question immediately after the examiner's reply.

Where the audio recorder stopped in the middle of students' utterances, that is shown by two slashes. Where students' utterances were accidentally interrupted by the examiner's reply and students were not able to complete their questions, I indicated that by putting two slashes in round brackets in places where the interruption occurred.

TABLE 6.1 Symbols used in the transcriptions

Italics:	Japanese words. e.g. *denshi renji* [*microwave oven*]
/ /:	Students' pronunciation of English words differed markedly from standard e.g. /microwave/
One dot:	A period of silence lasting roughly shorter than one second. e.g. Where is . a boy?
Three dots:	A period of silence lasting about more than one second. e.g. Where is … a boy?
[Inaudible]:	Students' inaudible or indecipherable utterances. e.g. Could you tell me … [inaudible] . mm ….
( ):	Examiner's utterances. e.g. Does the dog err…(*Do-zo nandemo iidesuyo* [= *Go on, you can ask any questions*]) … Does the dog eat cake?
//:	Audio recorder stopped in the middle of students' utterance. e.g. In your picture //
(//):	Students' utterance was interrupted by the examiner e.g. In your picture, there is a picture … on the wall (//)

6.2 Coding of question examples

6.2.1 Coding system

I used the transcriptions to examine each student's question examples individually, and identified which Q-form was appropriately applied, where I interpreted *appropriateness* in terms of *word order* required for each Q-form. As was also the case in previous research into the effects of instruction on stage advancement in interrogative acquisition, accuracy in producing questions, such as a mistake in number agreement, was disregarded if the required word order was present. Likewise, use of pragmatically inappropriate questions was disregarded in coding. I will give further explanation of these points in the next section.

My coding system was based on the six-stage developmental sequence for English question formation proposed by Pienemann and Johnston (1986) and on Spada & Lightbown's 1993 extension of the sequence (see also Pienemann, Johnston & Brindley, 1988), which has been discussed in Chapter 5. The codes used for nine Q-forms are shown in Table 6.2 with example questions on the next page. For Q-forms at stage 3 and lower levels, each code consists only of a stage level as follows: '1' (*One word/formula/sentence fragments*), '2' (*Canonical word order*) and '3' (*Fronting*). For Q-forms at stage 4 and higher, each code consists of a stage level and an abbreviation for the rules as in the following: '4wh' (*Pseudo-inversion*), '4yn' (*Y/N-inversion*), '5wh' (*Affirmative-Do/Aux-2nd*), '5Nwh' (*Negative-Aux/Do-2nd*), '6C' (*Cancel-inversion*), '6T' (*Q-tag*) and '6N' (*Y/N-Negative*).

I should remind the reader that in this study, I distinguished in stage 5 between the *affirmative* and *negative* form of *Aux/Do-2nd* for scrutiny, which have been conflated as one Q-form in previous research into the acquisition of question formation (e.g. Pienemann & Johnston, 1986: Pienemann, Johnston, & Brindley, 1988; Spada & Lightbown, 1993; 1999; Mackey, 1995; Mackey & Philp, 1998). These two stage 5 forms require the same degree of processing complexity (i.e. the placement of an auxiliary verb in the second position of the sentence before the subject). An *affirmative* form of *Aux/Do-2nd* can be applied as in 'How many classes do you take a week?' and its *negative* form as in 'Why haven't you done the dishes yet?'. I noticed from my teaching and learning experience that some learners fail to

TABLE 6.2 Developmental stages and examples of Q-forms

Developmental Stage		Code	Examples
Stage 1	***Single words, sentence fragments or formulae***	1	Why? Cat how many? Why not? What color the pendant?
Stage 2	***Canonical word order*** : Canonical word order with rising intonation.	2	There are two glasses on the table? You are looking for what shape of plates? There are two women, isn't it?
Stage 3	***Fronting*** : ***Do-fronting/Fronted-other*** Direct questions with main verbs and some form of fronting.	3	Does the man have a flower in his hand? How many flat mates are in your flat? How many flat mates you have? Where the cat is? Is there is a ball?
Stage 4	***Wh-inversion*** : In *Wh-/Q*-word questions the copula and the subject change positions.	4wh	Where is the cat? How many people are there in your picture?
	Y/N-inversion : In Y/N questions an auxiliary other than 'do' is in sentence-initial position.	4yn	Is there a cat in your picture? Have you sent an application form to the university? Can you tell me the name of newspaper?
Stage 5	***Affirmative-Aux/Do-2nd*** : An affirmative form of auxiliary verb or modal is placed in second position to a *wh-/Q-* word and before the subject followed by a main verb [applies only in main clauses/direct questions].	5wh	What can you see through the window? How many flat mate do you have? What is the man doing?
	Negative-Aux/Do-2nd : A negated form of auxiliary verb or modal is placed in second position to a *wh-/Q-* word and before the subject followed by a main verb [applies only in main clauses/direct questions].	5Nwh	Why can't we use the microwave? Why haven't you sent it yet?
Stage 6	***Cancel-inversion*** : Inversion is not present in subordinate clauses.	6C	Do you remember when Jill got married? When do you remember Jill got married? Could you tell me what time I can see you on Friday? Can you tell me if the woman is wearing earrings?
	Q-tag : An auxiliary and a pronoun are attached to the end of the main clause.	6T	In your picture there are two wine glasses, aren't there? The father doesn't live with his family, does he?
	Negative-Q : A negated form of Aux/Do is placed before the subject.	6N	Haven't you done the dishes yet?

develop *Negative-Aux/Do-2nd*, even after reaching the highest-level in the acquisition of question formation, and I wanted to investigate whether there were any differences in the acquisition of these two forms.

6.2.2 The procedures for coding question examples

To illustrate the procedures for coding question examples in detail, I will now cite from the transcripts actual examples for each Q-form.

Stage 1: ***One word/sentence fragments/formulae*** (Code: 1)

If students produced *single words*, *sentence fragments* or *formulae* as shown below, I coded examples as '1'.

*Why?
*What color the pendant?
Why not?

Stage 2: ***Canonical word order*** (Code: 2)

I coded examples as '2' if *canonical word order* was observed with rising intonation, as in the following two examples. If the same sentences were produced without rising intonation, I assigned 'NQ' (*non-question*) to such examples and excluded them from the analysis.

*There are two glasses on the table? ↑
*You are looking for what kind of plates? ↑

Stage 3: ***Fronting*** (Code: 3)

I coded examples as '3' (*Fronting*) where the auxiliary 'do', or other lexical elements such as a *wh-*/*Q*-word or an auxiliary verb other than 'do', were placed in sentence-initial position in direct questions[35]. For instance, in the first three examples below, the sentence was fronted by the auxiliary 'do', the auxiliary 'is', or the *Q*-word 'How many flat mates', respectively. I coded them as '3'.

Stage 3 questions can be target-like or non-target-like, and a mistake in number agreement was disregarded, because my purpose was not to examine *accuracy* but to see if the required *word order* for stage 3 was present.

*Do the man have a flower in his hand?
*Is the man have a flower in his hand?
*How many flatmates you have?

Stage 4: *Wh-inversion* (Code: 4wh)

For *Wh-inversion* to be considered as appropriately applied in *wh*-questions, the copula should be placed before the subject. In the first example below, this requirement was realised and I coded it as '4wh'. For the second example, although number agreement between the subject and the auxiliary was not correct, inversion was present . Stage 4 questions can be target-like or non-target-like, and provided the required word order (i.e. inversion) was present, I coded such examples as '4wh'. In the third example the student failed to invert the subject and the copula and I coded it as '3' (*Fronting*).

Where is the cat?
*How many people is there in your picture?
*How many people there are in your picture? → '3'

Stage 4: *Y/N-inversion* (Code: 4yn)

For *Y/N-inversion* to be regarded as appropriately applied in *yes/no*-questions, an affirmative form of auxiliary verb other than 'do' should be placed in sentence-initial position before the subject. In the first and second examples below, this requirement was realised and I coded them as '4yn'. A mistake in number agreement in the first example was disregarded, since stage 4 questions can be target-like or non-target-like. In the third example, inversion was not present and I coded it as '2', since it showed *Canonical word order* with rising intonation. If the same sentence was produced without rising intonation, I coded such examples as 'NQ' (*Non-question*) and excluded them from the analysis.

*Is there two woman in your picture?
Have you sent an application form to the university?
*You have sent an application form to the university? ↑ → '2'

Stage 5: ***Affirmative-Aux/Do-2nd*** (Code: 5wh)

As I noted earlier, I distinguished in stage 5 between the *affirmative* and *negative* form of *Aux/Do-2nd* for scrutiny. For *Affirmative-Ao/Aux-2nd* to be considered appropriately applied in *wh*/Q-word questions, an affirmative form of an auxiliary verb should be placed in second position to a *wh*/Q-word before a subject followed by a main verb. In the first and second examples below, this requirement was met and I coded them as '5wh'. Stage 5 questions can be target-like or non-target-like, so the third example was also coded as '5wh', although there was a mistake in number agreement.

What is the boy doing?
What can you see through the window?
*What is two lady doing on the sofa?

Stage 5: ***Negative-Aux/Do-2nd*** (Code: 5Nwh)

For *Negativ-Aux/Do-2nd* to be regarded as appropriately applied in *wh*-questions, a negated form of an auxiliary verb should be placed in second position after a *wh*/Q-word before a subject followed by a main verb. The two examples below show this word order and were coded as '5Nwh', although in the second example a learner made a mistake in number agreement.

Why hasn't the man cleaned the bathroom?
*Why haven't the man cleaned the bathroom?

Stage 6: ***Cancel-inversion*** (Code: 6C)

For *Cancel-inversion* to be considered as appropriately applied in indirect questions, there should be no inversion of the subject and the auxiliary/verb in the subordinate clause. In the first two examples below, this requirement was realised and I coded them as '6C'. In the third example, the student inverted the subject and the auxiliary verb in the subordinate clause and I coded it as '5wh' because the rule *Affirmative-Aux/Do-2nd* was observed. In addition, I assigned '4yn' to this example, since the student applied *Y/N-inversion* in the main clause as in 'Could you tell me' without any prompt to apply *Cancel-inversion*. We should recall that the examiner elicited examples of *Cancel-inversion* in the Spot-the-Difference task by giving the instruction 'Find more differences and start questions with *could*

you tell me...?'. If this example was produced after the examiner's prompt, I did not assign '4yn' to the main clause because the use of 'Could you tell me' was a repetition of what the examiner had said. In the transcription, I marked examples elicited by the examiner's prompt to apply a specific rule by an asterisk on the left of the given codes.

What time do you think I can see you?
Could you tell me what the man is doing?
*Could you tell me what is the man doing? → '4yn' + '5wh'

In coding indirect questions, I found that some students failed to insert a conjunction 'if/whether' or a *wh/Q*-word after a main clause and produced questions like the one below. Although inversion was not present in the subordinate clause, I did not code such questions as examples of *Cancel-inversion* because the placement of a conjunction 'if/whether' was part of the required word order for this rule. This question was produced without the examiner's prompt and I gave it a dual coding: '3 (*Fronting*) for the main clause and '2' (*Canonical word order*) for the subordinate clause.

*Do you know the cat is sleeping on the floor? → '3' + '2'

Stage 6: *Q-tag* (Code: 6T)

In coding question examples, I found that many learners who had not reached stage 6 tended to put the question tag 'isn't it?' irrespective of the subject and auxiliary/main verb of the sentence used in the main clause. This is known as 'suspect of formula' (Pienemann & Johnston, 1986:115). Many examples of *Q-tag* were elicited by the examiner's prompt at the end of the Spot-the-Difference task, in which the learners heard the sample question with the question tag 'isn't it?' and may have picked it up and used it as a chunk. For coding examples with a question tag, I set up the following criteria to make sure that those questions were productive examples of *Q-tag*.

To be considered to be a productive example of *Q-tag*,

(a) an affirmative form of an auxiliary verb should be present in the question tag if a negated form of an auxiliary verb is used in the main clause, and vice versa;

(b) there should be person agreement between the subject of the sentence and a pronoun in the question tag; and

(c) there should be number agreement between an auxiliary/main verb of the sentence and an auxiliary verb in the question tag.

For instance, the following three examples did not satisfy the criteria and were coded as '2' (*Canonical word order*).

*In your picture, err there is an apple on the TV, err is there? → '2'
*In your picture, the boy is sitting on the floor, isn't it? → '2'
*In your picture, there are two ladies, isn't there? → '2'

In the first example an affirmative form of auxiliary verb was used in both the main clause and the question tag. In the second, the pronoun in the question tag 'it' did not agree with the subject of the sentence 'the boy'. In the third, the number of the auxiliary in the question tag did not agree with that in the main clause. In fact, rising intonation is normally required for stage 2 questions but I made an exception because these examples were marked by the question tag as a question.

Thus, I applied the criteria to scrutinise the application of *Q-tag*, but stage 6 questions can be target-like or non-target-like, and a mistake in number agreement within the main clause was disregarded. In the first example below, the number of the auxiliary verb 'are' did not agree with that of the complement '24-hour shop' in the main clause, but this question was considered to be a productive example of *Q-tag*. We can see good evidence for the rule application of *Q-tag*, since the student held grammatical information about the subject and auxiliary verb in the main clause in short-term memory and used it to realise agreement when producing the question tag 'aren't there?'. For the same reason, the second example was also considered to be a case of the rule application of *Q-tag*. In this example, the student used the wrong form of the verb (present simple) both in the main clause and in the question tag. To describe the dog's motion drawn on the task sheet in the Spot-the-Difference task, the present progressive tense should have been used, but such mistakes were ignored as long as the required agreement set out in the criteria was realised.

*There are 24-hour shop near your house, aren't there?
In your picture, the dog eats a cake, doesn't it?

Another common mistake with a question tag was the attachment of a tag to interrogatives, as shown below. In such cases, I examined which Q-form was applied in the main clause and assigned a code accordingly and coded the example below as '4yn' (*Y/N-inversion*).

*In your picture, is there a cat, isn't it?' → '4yn'

Stage 6: *Y/N-Negative* (Code: 6N)

For *Y/N-inversion* to be regarded as appropriately applied in *yes/no*-questions, a negated form of auxiliary verb has to be placed in the initial-position of the sentence before the subject. In the first example below this requirement was realised and I coded it as '6N'. The second example was also considered a successful application of this rule, although it was produced on an occasion where the student was supposed to ask a question using *Negative-Aux/Do-2nd* as in 'Why haven't you sent an application form yet?' As discussed earlier, in coding I identified which Q-form was appropriately applied in question examples, where I interpreted *appropriateness* in terms of *word order* and use of pragmatically inappropriate questions was not taken into account.

Haven't you done the dishes yet?
Haven't you sent an application form yet?
*Haven't you? → '1'

In the third example, the student tried to apply this rule but the question was elliptic. Such elliptic questions are also used by native English speakers, but I decided that this example did not constitute evidence of productive usage of *Y/N-Negative* because there was a possibility that 'Haven't you?' was produced as a memorised chunk. This decision is supported by the fact that such elliptic questions were produced by only two students who had not reached stage 5.

I have shown the results of coding on the left of each example in the transcription (see Appendix 86 for the pretest, see Appendix 87 for the posttest, see Appendix 88 for the follow-up test). The coding system was straightforward and I did not have the question examples examined by other coders.

6.2.3 Coding of examples in which mistakes were self-corrected by student

In coding, I found that some students self-corrected mistakes in applying Q-forms. In principle, I assigned a code only to the corrected question and excluded previous attempts from the analysis. For instance, in the following example the student failed to apply *Affirmative-Aux/Do-2nd* at first and then corrected the mistake. I included only the corrected part of the utterance in the analysis and coded it as '5wh' (*Affirmative-Aux/Do-2nd*).

What the boy is doing . no, no. What is the boy doing? → '5wh'

In cases where students' self-corrections were triggered by the examiner's interruptions, I made exceptions. For instance, in the first example below, the student forgot to place the auxiliary verb 'can't' in the second place to the *wh*-word 'why' and failed to apply *Negative-Aux/Do-2nd* at first; the student then heard the examiner asking 'Pardon?' and corrected the mistake successfully. It is possible that the examiner's interruption drew the student's attention to the mistake, which eventually helped the student to correct it. In the second example, the student applied *Negative-Aux/Do-2nd* appropriately at first, but the examiner's interruption prompted the speaker to correct what the student had said, resulting in an incorrect application of this rule. In this context, the examiner did not mean to signal that there was a mistake by requesting repetition; however, arguably the student interpreted it not as a request for repetition, but as a signal of a mistake to be corrected in the preceding utterance. Learners tend to be sensitive to examiner's reaction to their utterances. In this example, the examiner's request might have caused a panic and urged the student to make changes to what she had said. In these cases, I decided not to code the corrected part and assigned a code only to the original utterance produced before the examiner's interruption. Before setting up each test, I had instructed the examiners not to ask the students to repeat what they had said, but sometimes such requests were made by mistake.

*Why the boy can't eat the cake? ↑ (Pardon?) Why can't the boy eat cake?

Why why . haven't you cleaned the bathroom? (Pardon? Sorry . why .?) *Why . why do you haven't cleaned the bathroom?

6.2.4 Other codes

In addition to the nine codes for Q-forms, I used the following five codes when I was not able to identify any Q-forms in examples: 'NQ', 'Skipped', 'Silence', 'No count' and 'No data'. 'NQ' refers to examples in which an utterance was a statement without rising intonation, as in 'The boy is sitting on the floor'. 'Skipped' indicates that students forgot to ask a question on the embedded occasions when they were supposed to elicit specific information from the examiner. 'Silence' indicates that students were silent either on the embedded occasions, when they were expected to elicit specific information from the examiner by asking questions, or on the obligatory occasions for applying specific Q-forms prompted by the examiner's instruction. Students' non-application of rule *x* on the obligatory occasion for applying rule *x* can be used as *negative* evidence for their acquisition of rule *x* and I used the contextual information shown by these two codes in the developmental stage assessment.

'No count' is used when identification of Q-form was not possible because either (a) students' utterances were interrupted by the examiner in the middle of questions or (b) the audio recorder stopped in the middle of students' utterances. 'No data' indicates that students' utterances were not recorded by mistake. Needless to say, examples assigned to one of these codes were not used as *negative* evidence for students' acquisition of Q-forms even if they were produced by the examiner's prompt to apply a specific Q-form.

6.3 The procedures for developmental stage assessment

By using the results obtained in the coding of question examples, I identified individual students' developmental stages on each test, using the question examples produced in the two conversation tasks (Guided Role-play and Spot-the-Difference). Following Mackey (1995, 1999) and Mackey & Philip (1998), which reported the effectiveness of intensive recasts on nontarget-like questions in one-to-one conversations in ESL learners' developmental stage gain, I employed two different assessment criteria: the one used for the posttest data was more rigorous than that for the pretest data. My intention was that this more rigorous criterion would allow a stronger claim that stage development had indeed occurred. I will now explain each criterion in detail in the next two sections.

6.3.1 The criterion used for stage assessment of the pretest

I used the following criterion for developmental stage assignment for the pretest.

> In order for students to be considered to have reached stage *x* as the highest stage, they had to produce:
>
> at least *two* different questions for this stage.

For example, in order for students to be considered to have reached stage 5 as the highest stage, they had to produce at least *two different* stage 5 questions. The combination of two questions might be: (a) two *Aux/Do-2nd* questions with different auxiliary verbs (e.g. 'How many classes do you take a week?'; 'What can you see through the window?'); (b) one *Aux/Do-2nd* question and one *Negative-Aux/Do-2nd* question; or (c) two *Negative-Aux/Do-2nd* questions with different auxiliary verbs. I should point out that learners' having reached stage 5 does not necessarily mean that they can produce examples for all rules at stage 5 and their non-production of examples for one rule at stage 5 does not constitute negative evidence in the developmental stage assessment. This is consistent with a view of acquisition as *emergence* rather than as *mastery*, proposed by Meisel, Clahsen, and Pienemann (1981). To illustrate

the procedures for the developmental stage assessment with actual data, I will show the process of coding I carried out for two students (Seigo and Yuya).

Seigo

Table 6.3 below shows the use of question examples of each Q-form by Seigo on the pretest. In the table, the occurrence of each Q-form is indicated. The figures in parentheses show the number of questions that were regarded as productive examples for each Q-form.

TABLE 6.3 Production of question examples of each Q-form (Seigo)

	Stage 1	Stage 2	Stage 3	Stage 4		Stage 5		Stage 6		
	1	*2*	*3*	*4wh*	*4yn*	*5wh*	*5Nwh*	*6C*	*6T*	*6N*
Pretest	3 (2)	0 (0)	13 (7)	3 (2)	12 (4)	6 (4)	2 (2)	0 (0)	3 (3)	1 (1)

For instance, he produced three tag questions and one *Y/N-Negative* question as the highest stage examples, and I examined if they were productive examples of each Q-form. In the three tag questions cited below, Seigo used three different auxiliary verbs in the tag ('are there?', 'aren't there?', 'doesn't she?') with the required agreement and these were considered productive examples of *Q-tag*.

> In your picture, err aren't there . err one cat . uh. In your picture, there are not there are not cats on the table, are you uh are there? ↑
> In your picture, err there are three trees in the garden, aren't there? ↑
> In in your picture, umm . the uh its girl wear yellow shirt, err err don't . doesn't she? ↑

For only one example of *Y/N-inversion* quoted below, there was a chance that the inversion of the negated form of copula and the subject in this question ('Aren't there') was a formulaic token, but as there were four different stage 6 questions in total in Seigo's sample, I concluded that he had reached stage 6.

> Aren't there 24-hour shop in uh near your flat? ↑

Yuya

Yuya produced two *Q-tag* questions as the highest stage examples (see Table 6.4 below).

TABLE 6.4 Production of question examples of each Q-form (Yuya)

	Stage 1	Stage 2	Stage 3	Stage 4		Stage 5		Stage 6		
	1	*2*	*3*	*4wh*	*4yn*	*5wh*	*5Nwh*	*6C*	*6T*	*6N*
Pretest	1 (1)	3 (3)	18 (10)	2 (1)	10 (3)	6 (4)	1 (1)	0 (0)	2 (1)	0 (0)

However, in both examples Yuya used the same auxiliary verb 'isn't' in the tag (see below) and one of them was invalidated as evidence, although he realised the other requirement for tag questions, such as person and number agreement, discussed in Section 6.2.2.

In your picture, err . is err there is a uh badminton racket on the floor, isn't there? ↑
In your picture, err it is err eleven o'clock now, isn't it? ↑

A closer examination of Yuya's other question examples has also shown that he failed to apply *Q-tag* once and *Cancel-inversion* twice on obligatory occasions for each rule, as shown below.

Misapplication of *Q-tag*

*In your picture, err . there is err . is there uh there is no uh there is only three glasses and the err newspaper . on the table, . isn't they? ↑

Misapplication of *Cancel-inversion*

*Could you err could tell me err how many err trees outside err of the err window?
*Could you me err could you tell me that is there err soccer ball under the table?

In the light of these observations, I concluded that Yuya had not reached stage 6. I continued to examine his other question examples at the next highest stage (i.e. stage 5) and found that he produced six *Affirmative-Aux//Do-2nd* questions in which he used four different auxiliary verbs (does, do, can, will)

and one *Negative-Aux/Do-2nd* question (see below). The production of these stage 5 questions was therefore sufficient evidence for his acquisition of stage 5.

Affirmative-Aux/Do-2nd

How much the does it cost?
Which which bank do you recommend?
When will my classmates come here? Uh. What time will my classmates come here?
What time can I can you see me? ↑
What does he err drink?
What the er lady near the window . What the . What does err the lady near the table doing mm do now?

Negative-Aux/Do-2nd

Why why can't you Why can't I why can't I use the washing machine at the moment?

I carried out similar developmental stage assessments for each student. While identifying individuals' developmental stage for the pretest, I produced an implicational table for each student to make sure that my stage assignment was appropriate. Table 6.5 below shows the implicational scales created for Seigo and Yuya. A'+' indicates that the students produced at least two different questions for a given stage; a '-' indicates that they produced only one or no example for a given stage; '/' indicates that there were no occasions for applying rules. It should be noted that although Yuya produced only one stage 1 question, he gained a '+' for stage 1 because it is apparent that he could apply a stage 1 rule (e.g. *One word question*). We can see from their implicational scales that, for both students, the acquisition of the highest level entailed the acquisition of all lower levels and they did not skip any stage.

TABLE 6.5 Stage assignment (Seigo and Yuya)

		Stage 1	Stage 2	Stage 3	Stage 4	Stage 5	Stage 6	Assigned stage
Seigo	Pretest	+	/	+	+	+	+	Stage 6
Yuya	Pretest	+	+	+	+	+	-	Stage 5

The requirement used for stage assignment for the pretest is essentially arbitrary, but it is the criterion employed by Mackey (1995, 1999) and Mackey and Philp (1998), and Spada and Lightbown (1993;

1999) for their stage assessment on the pretest and represents a 'more conservative measure' than a 'one usage measure'. However, as discussed in Section 3.3.3, the use of an emergence-based criterion entails the difficulty of distinguishing a formula (i.e. an unanalysed chunk) from an example of successful application of a Q-form. For instance, during coding I frequently observed formulaic constructions such as the question tag 'isn't it?', 'Do you have..?', 'What's...?', 'Is there..?', and 'Is it..?'. To minimise such misjudgement in analysing individual's gain in developmental stage, I employed a stricter criterion for stage assignment for each posttest in order to allow a stronger claim that stage advancement and the acquisition of the target rules had indeed occurred.

6.3.2 The criterion used for stage assessment of the posttests

I used the following criterion for developmental stage assignment for the two posttests and for overall stage gain assessment through the treatment.

> In order for students to be considered to have reached at stage *x* as the highest stage, they had to produce:
>
> (a) at least *two* productive usages
> (b) of *two different* Q-forms for the given stage.
>
> In addition, for students to be considered to have gained a stage in relation to the pretest, they had to achieve:
>
> (c) a higher stage than the pretest on *both* posttests.

Although this requirement is also essentially arbitrary, it represents a 'more conservative measure than the 'two productive usages' at one time measure used in previous studies (Pienemann & Johnston, 1987; Pienemann, Johnston & Brindley, 1988; Spada & Lightbown, 1993, 1999). This is because it entails two usages of two forms on more than one occasion' (Mackey & Philp, 1998:347).

The two kinds of assessment criteria used in this study were employed by Mackey (1995) and Mackey

and Philp (1998), in which they reported that immediate, intensive recasts on ESL learners' nontarget-like interrogatives in one-to-one conversations were effective in helping them advance in developmental stage in the acquisition of question formation. In these studies, the students were given three posttests within four weeks of the last treatment session and they were considered to have gained a stage if they satisfied the requirement on any combination of two of the three post tests given within one month. In other words, students were considered to have made advancement in developmental stage only when their gain was sustained for either two weeks or four weeks.

To illustrate the procedures for individual's stage gain assessment, I will now show how I carried out developmental stage assignment for two students (Yo and Jo).

Yo

On the pretest, Yo produced three productive examples of stage 6 questions as the highest stage examples: one *Cancel-Inversion* question and two *Q-tag* questions with two different auxiliary verbs (see below). The criterion used for stage assignment for the pretest required 'at least *two* different questions' for the given stage and these observations constituted sufficient evidence for assigning Yo to stage 6.

Cancel-Inversion

Could you tell me . what is . under the table?

Q-tag

In this picture uh in yours picture, the tennis racket is . uh costs . one hundred and seventy . dollars, doesn't it? ↑

I think, . guest room is nea- guest room was included, wasn't it? ↑

I then examined Yo's question examples produced on the posttest, using the stricter criterion. Table 6.6 below is an extract from the summary of the number of each Q-form examples produced by each student in each test (see Appendix 85) and the result of stage assignment for each test. In this table, the occurrence of each rule on each test is indicated in each column and the figures in parentheses show the number of productive question examples used as evidence for the acquisition of each rule.

TABLE 6.6 Production of question examples of each Q-form (Yo)

	Stage 1	Stage 2	Stage 3	Stage 4		Stage 5		Stage 6			Assigned stage
	1	*2*	*3*	*4wh*	*4yn*	*5wh*	*5Nwh*	*6C*	*6T*	*6N*	
Pretest	1 (1)	1 (1)	8 (5)	1 (1)	11 (3)	9 (6)	1 (1)	2 (1)	2 (2)	0 (0)	6
Posttest	4 (3)	2 (2)	14 (8)	3 (1)	8 (2)	9 (5)	1 (1)	1 (1)	1 (1)	1 (1)	3
Follow-up test	5 (4)	1 (1)	15 (12)	6 (1)	7 (3)	7 (6)	0 (0)	1 (1)	3 (1)	0 (0)	3

On the posttest, Yo produced one example for three stage 6 Q-forms (see below).

Could you tell me what is on the television?
In your picture, err circled table is between . chair. and television, isn't it?
Isn't there a . restaurant near home?

If I had applied the criterion used for stage assignment for the pretest, Yo would have been assigned to stage 6 again, but I employed the stricter criterion for the two posttests and he was not assigned to stage 6 for the posttest, since it required students to produce 'at least *two* productive usages of *two different* Q-forms' for the given stage.

I then examined his stage 5 question examples. Table 6.6 above shows that he produced nine *Affirmative-Aux/Do-2nd* and one *Negtive-Aux/Do-2nd* questions. In the nine *Affirmative-Aux/Do-2nd* questions (see below), he used five different auxiliary verbs ('will', 'do', 'can', does', 'is') and they were considered productive examples of this Q-form.

When will your friend come here?
What kind of sports did do they like especially?
Why can you . why can you eat cake?
Where does he live?
Could you tell me . what . the . what is the young guy doing?
Which bank do you recommend?
How many classes do you take in a day?
What color do the guy wear?
When mh when can I meet you at Tuesday?

For the other stage 5 rule, *Negative-Aux/Do-2nd*, there was only one example of successful application of this rule in Yo's posttest data (see the first example below), while he failed to apply this rule in the presence of two more obligatory occasions to apply this rule, which served as evidence for his non-acquisition of this form. Although there were five productive examples of *Affirmative-Aux/Do-2nd*, I concluded that Yo had not reached stage 5 and examined his stage 4 question examples to identify his developmental stage, repeating the same procedures.

Why didn't you sen- send . /an/ application form for the scholarship?
*Why err the man haven't bought skis?
*Why can you . why can you eat cake?

On the posttest, Yo produced three *Wh-Inversion* questions and eight *Y/N-Inversion* questions (see Table 6.6). However, he used the same copula 'is' in the three *Wh-Inversion* questions and two examples were invalidated. In the eight *Y/N-Inversion* questions he used the same auxiliary 'is' in seven examples and six examples were invalidated as evidence. These observations did not constitute sufficient evidence for assigning him to stage 4. I eventually assigned Yo to stage 3, since there were eight stage questions with different fronted elements. It should be noted that, in order for students to be assigned to stage 3 for the two posttests, four different fronted elements should be present in stage 3 question examples. For the two posttests, no student was assigned to stage 2 or stage 1.

The next step for Yo's stage gain assessment was to identify his developmental stage for the follow-up test. Table 6.6 shows that there was only one example of successful application for two stage rules. I also found in his follow-up test data that he tried to apply *Cancel-Inversion* on other four occasions, but succeeded only once (see the first example below).

Successful application of *Cancel-Inversion*
Could you tell me what is on the mountain . top of mountain?

Misapplication of *Cancel-Inversion*
*Could you tell me where is the basket?
*Could you tell me where is he?
*Could you tell me is it sure?

Yo also produced three *Q-tag* questions (see below), he used the same auxiliary verb 'isn't' in the tag

and two of the three examples were invalidated. Moreover, a closer look at his data showed that he failed to apply *Q-tag* twice on this test.

In your picture, tulip is drawn err near the window, isn't it? ↑
In your picture, the man . is having a ph- phone . not a flower, mm . isn't he? ↑
The dormitory is mh including a computer room, isn't it? ↑
*In your picture, mm the umbrella cost twelve dollars, . isn't it? ↑
*Isn't isn't there supermarket is near the dormitory, is it? ↑

These observations did not constitute sufficient evidence for assigning him to stage 6. I then examined Yo's stage 5 question examples. He produced six productive examples of *Affirmative-Aux/Do-2nd* (see Table 6.6), but there was only one *Negative-Aux/Do-2nd* question, and I did not assign Yo to stage 5. Likewise, I did not assign him to stage 4 because of the lack of sufficient evidence for this stage assignment. I eventually assigned Yo to stage 3 again for the follow-up test. As a result, he was considered to have regressed by three stages in developmental stage on the posttests.

As I have stressed, the assessment criterion used for stage assignment was more rigorous than that used for the pretest and, in order for many students who had been assigned to stage 6 on the pretest to be assigned to stage 6 on the two posttests, they had to produce more stage 6 questions on the two posttests - at least two productive examples for two different stage 6 rules. In other words, if the treatment is not effective in helping students acquire higher-level rules, many students can show regression in developmental stage.

I will show one more example of stage gain assessment for another student (Jo), who was assigned to stage 5 on the pretest and was considered to have advanced to stage 6 on the posttests.

Jo

On the pretest, Jo produced only one stage 6 question as shown in Table 6.7 below. In this example he applied *Q-tag* successfully (see the first example below), but failed to do so in other two examples.

In your picture, the newspaper's name is *Daily Times*, isn't it?
*In your picture, the woman in the kitchen cooking the soup, isn't it? ↑
*In your picture, the cat sleeping on the table, isn't it?

Jo also failed to apply *Cancel-Inversion* three times on the pretest (see below) and there was no example of successful application of this rule.

*Could you tell me the cat doing?
*Could you tell me the woman doing?
*Could you tell me how many trees do you have on the picture? ↑

These observations show that Jo had not acquired these two stage 6 rules. I then examined his stage 5 rules and there were three productive examples of this rule on the pretest, which is sufficient evidence for assigning Jo to stage 5.

TABLE 6.7 Production of question examples of each Q-form (Jo)

	Stage 1	Stage 2	Stage 3	Stage 4		Stage 5		Stage 6			Assigned stage
	1	*2*	*3*	*4wh*	*4yn*	*5wh*	*5Nwh*	*6C*	*6T*	*6N*	
Pretest	9 (8)	4 (4)	17 (7)	0 (0)	3 (2)	7 (3)	0 (0)	0 (0)	1 (1)	0 (0)	5
Posttest	1 (1)	3 (3)	12 (7)	1 (1)	10 (4)	3 (2)	1 (1)	6 (5)	5 (2)	0 (0)	6
Follow-up test	2 (2)	2 (2)	13 (7)	4 (1)	4 (4)	6 (4)	1 (1)	5 (4)	5 (4)	0 (0)	6

On the posttest, we can see from the table above that Jo produced more stage 6 questions than on the pretes: six *Cancel-inversion* questions and five *Q-tag* questions, and I looked for at least two productive questions for each Q-form. In the six *Cancel-inversion* questions (see below), Jo used two different auxiliary verbs as in 'can meet' and 'is reading', two kinds of copula as in 'what time it is' and 'if there are', and one main verb without an auxiliary as in 'what kind of facilities you have', which was sufficient evidence for his acquisition of this rule.

Could you tell me host father err . how many children the host father have?
Could you tell me what kind of facilities . you have in your kitchen?
Could you tell me the time what time . err can uh we can meet in the host family . on Tuesday?
Could you tell me what kind of newspaper the man reading the man is reading?
Could you tell me what time in the what time is . what time . it is?
Could you tell me if there are an apples an apples on the TV?

In the five tag questions quoted below, Jo used two different auxiliary verbs in the tag as in 'isn't it?' and 'is he?', which was also sufficient evidence for his acquisition of this rule. Given these observations, I concluded that Jo had reached stage 6 on the immediate posttest and then continued to identify his developmental stage on the follow-up test.

The and the 24-hour shop is near the house, isn't it?
Your host father is . not living with his family, err is he?
The house is in the . near by the language center, err isn't it?
The restaurant is not . near the . the house, isn't is it?
In your picture, the dog . dog is eating cookies, isn't it?

On the follow-up test, Jo produced five examples for two stage 6 rules. Four of the five *Cancel-inversion* questions (see below) were considered productive examples of this rule, since in the first two examples he used two different auxiliary verbs as in 'I can pay' and 'the girl is wearing' in the subordinate clause, one main verb without an auxiliary as in 'how many classes you have', and one kind of copula 'be' as in 'where the basket is'.

Could you tell me err if I can err pay by credit card?
Could you tell me if the girl is wearing glasses?
Could you tell me how many classes you have in a week?
Could you tell me where the basket is?
Could you tell me where the two glasses is? ( ) Could you tell me where is . the glass uh where the glass is?

Moreover, in the five tag questions (see below), Jo used two different auxiliary verbs as in 'isn't he?' and 'doesn't it?', which was sufficient evidence for his acquisition of *Q-tag*.

I think, . the dormitory have the language uh is nearby the language center, isn't it?
The dormitory have the computer room, doesn't it . uh . doesn't it?↑
The manager of your dormitory is living in your buildings, isn't he?↑
In your picture, the . err. In your picture, the dog is between the man and woman, isn't he?↑
In your picture, the girl is wearing yellow shirts, isn't isn't she?

Thus, he produced two or more productive questions for two different stage 6 rules on the follow-up test, so I assigned him again to stage 6. Jo sustained his gain for seven weeks and I concluded that he had advanced in developmental stage from 5 to 6.

For the developmental stage gain assessment, I had question examples produced by 18 of the 60 students on the three tests examined by a native English speaker and compared her judgments with mine. She used the results of the coding of question examples obtained in Section 6.2 and the same assessment criteria to examine individuals' developmental stage on each test. For the pre- and posttests, there was unanimity on all 18 students' stage assignment. For the follow-up test, the two coders agreed on 17 students' stage assignment , representing inter-coder reliability of 94 percent.

6.3.3 Justification for using the emergence-based criteria for stage gain assessment

In the previous section, I demonstrated that the use of a stricter criterion for the two posttests required students to produce more question examples at higher stage(s) on both posttests than on the pretest, and that many question examples were invalidated as evidence when students used the same auxiliary verb in questions with the given Q-form. This is the criterion employed by Mackey (1995, 1999) and Mackey and Philp (1998) and, to the best of my knowledge, this is most rigorous among emergence-based criteria used in previous research into interrogative acquisition.

Having said that I employed the same requirement used in Mackey (1995, 1999) and Mackey and Philp (1998), I should point out the actual requirement in this study was in fact even more rigorous than that used in the previous studies. For instance, in my study, in order for students to be considered to have gained a stage, they were required to sustain their gain over a period of 6-7 weeks. In the previous three studies, the students were given three posttests within four weeks of the last treatment session and they were considered to have gained a stage if they satisfied the same requirement on any combination of two of the three post tests given within one month. For example, the first posttest was given the day after the last treatment session, the second administered one week after the first posttest, and the last given four weeks after the first posttest. In contrast, in this study the posttest was given one week after the last treatment session and the follow-up test was administered some 6-7 weeks after the last treatment. In other words, the definition of the term 'sustained effects of the treatment' in my study is 'learners' sustained gain over either six or seven weeks', while in the previous studies it means 'learners'

sustained gain of either two weeks or four weeks', which was a third or little more than half the length of period of time when compared to this study.

Given my use of the more demanding requirement for stage assignment for the two posttests and for sustained gain assessment in this study, it seems reasonable to make a claim that there is a good chance that stage advancement in interrogative acquisition - emergence of higher level Q-forms in learners' interlanguage - had indeed occurred. However, I should exercise caution in drawing conclusions, since the use of an emergence-based criterion entails the difficulty of distinguishing a formula from an example of successful application of a Q-form.

Sections 6.2 and 6.3 have provided an overview of the coding system and the procedures with, I believe, enough clarity for replication studies to be carried out, both for the developmental stage assessment for the pretest and for the stage gain assessment on the two posttests. In the next section I will present my analysis of sustained gain in developmental stage after the treatment.

6.4 Findings and discussion

6.4.1 Individuals' developmental stage on the pretest

Table 6.7 below summarises the results of my developmental stage assessment for the pretest. The number of students assigned to stage 6 was 22 in the Treatment group (TG) and nine in the Comparison group (CG). For Stage 5, the number of students assigned to this stage was nine in the TG and three in the CG. For the rest of the students assigned to Stage 4 or below, all were in the TG and the number of students assigned to each stage was four at Stage 4, 12 at Stage 3, and one at Stage 1.

TABLE 6.8 Results of developmental stage assessment (pretest)

Developmental stage on the pretest	TG (N= 48)	CG (N= 12)	TOTAL (N=60)
Stage 6	22	9	31
Stage 5	9	3	12
Stage 4	4	0	4
Stage 3	12	0	12
Stage 2	0	0	0
Stage 1	1	0	1

In the process of developmental stage assessment, based on the results of the use of Q-forms for each student on the pretest (see Appendix 85), I produced an implicational scale for all 60 students in Table 6.8 on the next page. For each student, the implicational relationship was not contradicted by negative evidence (i.e. marked by a '-') in the hierarchy in the presence of positive evidence (marked by '+') for any higher levels. In other words, no student skipped a stage in their acquisition of question formation, which provides strong support for the hypothesised implicational relationship in the acquisition of question formation. These results also give support to my analysis of the developmental stage assessment for the students.

TABLE 6.9 Implicational table (pretest)

	Stage 6	Stage 5	Stage 4	Stage 3	Stage 2	Stage 1
Yo	+	+	+	+	+	+
Dan	+	+	+	+	+	+
Goro	+	+	+	+	+	+
Jiro	+	+	+	+	+	+
Taro	+	+	+	+	+	+
Kiko	+	+	+	+	+	+
Miki	+	+	+	+	+	+
Rose	+	+	+	+	+	+
Keiko	+	+	+	+	+	+
Moto	+	+	+	+	+	+
Mat	+	+	+	+	+	+
Jin	+	+	+	+	+	+
Hide	+	+	+	+	+	+
Go	+	+	+	+	+	+
Toby	+	+	+	+	+	+
Waka	+	+	+	+	+	+
Mama	+	+	+	+	+	+
Fuku	+	+	+	+	+	+
Riko	+	+	+	+	+	+
Hana	+	+	+	+	+	+
Tomi	+	+	+	+	+	+
Koara	+	+	+	+	+	+
Junko	+	+	+	+	+	+
Hiro	+	+	+	+	+	+
Seigo	+	+	+	+	+	+
Ru	+	+	+	+	+	+
Yuta	+	+	+	+	+	+
Yan	+	+	+	+	+	+
Ken	+	+	+	+	+	+
Tami	+	+	+	+	+	+
Saya	+	+	+	+	+	+
Sae	'	+	+	+	+	+
Mi	'	+	+	+	+	+
Seto	'	+	+	+	+	+
Shin	'	+	+	+	+	+
Yuya	'	+	+	+	+	+
Taka	'	+	+	+	+	+
Kaji	'	+	+	+	+	+
Nae	'	+	+	+	+	+
Tana	'	+	+	+	+	+
Jo	'	+	+	+	+	+
Nobu	'	+	+	+	+	+
Toyo	'	+	+	+	+	+
Mo	'	'	+	+	+	+
Kako	'	'	+	+	+	+
Rika	'	'	+	+	+	+
Omi	'	'	+	+	+	+
Nori	'	'	'	+	+	+
Tomo	'	'	'	+	+	+
Ai	'	'	'	+	+	+
Saki	'	'	'	+	+	+
Sato	'	'	'	+	+	+
Ko	'	'	'	+	+	+
Chie	'	'	'	+	+	+
Maya	'	'	'	+	+	+
Yaya	'	'	'	+	+	+
Sue	'	'	'	+	+	+
Yuko	'	'	'	+	+	+
Eri	'	'	'	+	+	+
Tsuta	'	'	/	'	'	+

6.4.2 Group assignment

In order to examine the effects of the treatment on individuals' gain in developmental stage in relation to their readiness to learn stage 6 Q-forms, as shown in Table 6.9 on the previous page, I classified students as *Stage6-Ready* or *Stage6-Unready*. I assigned 12 students who had reached stage 5 on the pretest (TG=9, CG=3) to the *Stage6-Ready* group because they were ready to learn stage 6 Q-forms which were at the next stage of their development. I assigned 16 students (all in the TG) who had reached stage 4 or below on the pretest to the *Stage6-Unready* group, since they first needed to acquire Q-forms at stage 5 (and at lower stages for some students, depending on their current developmental stage) before learning Q-forms at stage 6. These 16 *Stage6-Unready* students comprised four students at stage 4, 11 students at stage 3 and one student at stage 1. It should be noted that one student's data (Ko at stage 3) was excluded from this group assignment, since her posttest performance in the Spot-the-Difference task was not recorded for technical reasons and examination of her stage development on the posttests was not possible. In the CG there were no *Stage6-Unready* students.

TABLE 6.10 Group assignment

Group	Developmental stage on the pretest	Treatment group	Comparison group
Stage6-Ready	Stage 5	9	3
Stage6-Unready	Stage 4	4	0
	Stage 3	11	0
	Stage 1	1	0

6.4.3 Sustained effects over 6-7 weeks on individuals' gain in developmental stage

I have summarised the results of stage assignment for each student for each test and the results of stage gain assessment in Table 6.10 below.

TABLE 6.11 Results of developmental stage assessment on the three tests

Treatment Group					Comparison Group				
	Pretest	Posttest	Follow-up test	Showed a sustained gain?		Pretest	Posttest	Follow-up test	Showed a sustained gain?
Stage-6 Ready (N=9)					*Stage-6 Ready* (N=3)				
Jo	5	6	6	Yes	Seto	5	4	4	No
Nobu	5	6	6	Yes	Mi	5	3	5	No
Tana	5	6	6	Yes	Sae	5	3	3	No
Nae	5	6	6	Yes					
Kaji	5	6	6	Yes					
Toyo	5	3	6	No					
Yuya	5	5	6	No					
Shin	5	3	3	No					
Taka	5	5	3	No					
Stage6-Unready (N=16)					*Stage6-Unready* (N=0)				
Mo	4	6	5	Yes					
Omi	4	3	3	No					
Rika	4	4	3	No					
Kako	4	3	3	No					
Eri	3	3	3	No					
Yuko	3	3	3	No					
Sue	3	3	3	No					
Yaya	3	3	3	No					
Maya	3	4	3	No					
Chie	3	3	3	No					
Sato	3	3	3	No					
Saki	3	3	3	No					
Ai	3	3	3	No					
Tomo	3	3	3	No					
Nori	3	3	3	No					
Tsuta	1	3	3	Yes					

6.4.3.1 Stage6-Ready

In the TG five of the nine Stage6-Ready students (56%), all at the university, gained a stage and moved up to stage 6. This means that these five students produced at least two different productive questions for two different stage 6 Q-forms, not only on the immediate posttest but also on the follow-up test. In other words, they sustained their gains for some seven weeks after the treatment. In contrast, in the CG none of the three *Stage-Ready* students gained a stage.

Of the other four students who did not gain a stage in the TG, no student showed an immediate gain on the posttest; although they produced more stage 6 questions and two (Taka, Yuya) were assigned to stage 5 again and the other (Toyo, Shin) were assigned to stage 3. However, two of the four (Toyo, Yuya) increased by a stage to stage 6 on the follow-up test, suggesting that the treatment had a delayed effect. In the CG no such positive change was observed in any of the three students.

Given that restructuring of interlanguage can take time, it would not be surprising if some students did not gain a stage on the immediate posttest. If we total the number of students showing the sustained and delayed gain, the effects of the treatment were found in eight of the nine Stage6-Ready students in the TG (89%), while none of the three Stage6-Ready students in the CG showed any gain at all in developmental stage. The difference between the two conditions shows that the sustained gain observed in the TG was brought about by the treatment rather than by task repetition in the tests and that the treatment was effective in helping the five students advance in developmental stage. The number of students in each condition was of course small and we should be cautious about drawing a firm conclusion, but these results do underscore the crucial importance of providing instructions to help learners advance in developmental stage in the acquisition of question formation.

6.4.3.2 Stage6-Unready

With regard to the Stage6-Unready group, only two of the 16 students showed a sustained gain: one (Mo) advanced from stage 4 to stage 6 on the posttest and decreased by a stage to stage 5 on the follow-up test, and the other (Tsuta) advanced from stage 1 to stage 3 on the posttest and sustained the

gain until the follow-up test. However, the other 14 students (88%) did not show any sustained or delayed gain after the treatment. This shows that the treatment focusing mainly on stage 6 Q-forms was not beneficial in helping most students who had not reached stage 5 advance in developmental stage. These results show that learners' readiness to learn higher-level rules played a crucial role for advancement in developmental stage, supporting the Teachability Hypothesis, which proposes that the learner can acquire structures at the next stage of their development.

It was hoped that inclusion of one of the two stage 5 forms (i.e. *Negative-Aux/Do-2nd*) as a target rule would help students ready to learn this rule to increase in developmental stage, but the fact that only one of the four students at stage 4 on the pretest (Mo) gained a stage shows that most students did not benefit. One possible reason is that, in the treatment, this stage 5 rule was taught together with stage 6 *Y/N-Negative* in the last treatment session, after the students had learnt two stage 6 forms (*Q-tag* and *Cancel-inversion*) in the first three classes. It could be that the earlier input focusing on a stage beyond their next stage of development (i.e. stage 6 rules) confused many students at stage 4 since they had to acquire stage 5 rules first.

I will now compare these findings with those in the aforementioned study by Mackey and Philp (1998) in which they employed the same two kinds of criteria for the assessment of learners' gain in developmental stage in interrogative acquisition. In their study, they examined the effect of intensive recasts on the non-target-like questions at stage 3 and above given by a trained native speaker (NS) in the meaning-focused information-gap tasks, and compared the results with those found in learners who received the same treatment without intensive recasts. The number of learners in each condition was nine and six, respectively, and their developmental stage on the pretest was between 3 and 5. The tasks were designed to provide plentiful opportunities to use many questions between stage 2 and 6, so that the they would produce questions at higher level(s) than their current stage. The instruction given to both conditions did not provide explicit teaching of grammar and the NS did not correct the learners' mistakes during the tasks. In other words, the learners in the treatment group received typical reactive FonF instruction.

After receiving five sessions (15-25 minutes/session/day) on five consecutive days, seven of the nine

learners who received intensive recasts (78%) showed a sustained gain over either two or four weeks, while in the other condition only one of the six learners (17%) showed a sustained gain, despite their readiness. Mackey and Philp (1998) concluded that interaction-rich instruction with intensive recasts was more beneficial than interaction alone in facilitating an increase in production of targeted higher-level Q-forms, and therefore for interlanguage development. Indeed, these results show the effectiveness of this type of reactive FonF instruction, but I should point out that in their study none of the learners at stage 5 in both conditions advanced to stage 6. Mackey (1999) explained that this might be due to the fact that the production of stage 6 questions was much less frequent than of those at lower stages. This suggests the limitation of this type of reactive FonF instruction to provide opportunities to ask questions with stage 6 rules and to help stage 5 learners advance in developmental stage.

In my study, the proportion of students showing the sustained gain from stage 5 to stage 6 (5 of 9=56%) was slightly lower than that observed in Mackey and Philp's 1998 study (7 of 9=78%). However, given that the period of time over which the students were required to sustain gains in this study was seven weeks - twice or three times as long as that required in Mackey and Philp (1998) – it is remarkable that the proactive form-focused instruction had such long-term effects.

Furthermore, in this study delayed effects of the treatment have been observed in two of the four stage 5 students. In the treatment, the students did not receive immediate feedback such as recasts from the teacher; I had been concerned as to how many students would benefit from the instruction without immediate teacher feedback on individuals' mistakes, such as recasts, since the treatment consisted of explicit focus on the target forms and the noticing tasks in which I encouraged my students to give feedback on each other's mistakes to compensate for the lack of immediate feedback from me. Given that the same criterion was used in both studies for stage assignment and that the requirement for the length of sustained gain was longer in this study, it is striking that the proactive form-focused instruction in this study was as effective as the recast-rich, reactive FonF instruction provided in Mackey and Philp (1998) to help learners advance in developmental stage, even without immediate, intensive teacher feedback.

These findings are encouraging for teachers, especially for those who teach large-size classes where

intensive teacher feedback is difficult to provide. In this study, in order to compensate for the lack of teacher feedback during pair work, I employed the following three means. First, I used the classroom tasks designed to provide learners with opportunities for *noticing gaps* between what they had heard in the mini-lecture on the target forms, and what they could actually do in using them. Second, I encouraged the students to take an active role in giving *peer feedback* on their partner's mistakes in pair work by showing the video of modelling which introduced effective ways of doing that in the conversation task which the students were going to be given. Third, I gave the whole class *summary feedback* on common mistakes in using the target Q-forms in the middle of the conversation task in each session - a commonly used technique.

In this study, however, it is not possible to show exactly which component of the treatment contributed to the sustained development. Equally, I cannot show from the available data what kinds of peer feedback took place in the noticing tasks in the treatment, and whether or not such feedback contributed to the long-term development, although this does not necessarily weaken the results of the study. One way to examine the relationship between peer feedback and L2 development would be to analyse the students' conversations in pair work recorded in the treatment, but that was beyond the scope of this study. In Chapter 8, as an alternative means of exploring the benefits of peer feedback in the conversation tasks, I will show my analysis of the questionnaires given after each treatment session, in which I asked the students for their perceptions of the usefulness of peer feedback in the classroom tasks designed to assist noticing. This will provide insight into both the possibilities and the limitations of on-line peer feedback for English interrogative acquisition.

6.4.4 Effects of the treatment on students at stage 6

I will now examine changes in the 31 students who had reached the highest level, stage 6, on the pretest and were excluded from the stage gain analysis in the previous section. These 31 students (TG=22; CG=9), all at the university, were assigned to stage 6 on the pretest because they produced two different stage 6 questions, but only eight of the 31 students produced two or more productive examples for two different stage 6 rules, which will be discussed in detail later in this section. In the stage assessment for the two posttests, I applied the stricter requirement and in order for many of the 31 students to be

considered to have maintained stage 6, many students had to produce more stage 6 questions on each posttest than on the pretest.

Table 6.12 below summarises the results of stage assignment for the 31 stage 6 students. In the TG 12 of the 22 students (55%) were assigned to stage 6 on both posttests and were considered to have maintained stage 6 (marked by 'Yes'). On the other hand, in the CG only one of the nine students (11%) was considered to have maintained stage 6, and that ratio is smaller than that of the TG. This suggests that the treatment was effective in helping learners produce more stage 6 questions for two different Q-forms.

TABLE 6.12 Results of developmental stage assessment for the students at stage 6 on the pretest

Treatment Group					Comparison Group				
	Pretest	Posttest	Follow-up test	Maintained stage 6?		Pretest	Posttest	Follow-up test	Maintained stage 6?
Stage-6 students (N=22)					*Stage-6 students* (N=9)				
Saya	6	6	6	Yes	Taro	6	6	6	Yes
Yan	6	6	6	Yes					
Yuta	6	6	6	Yes					
Seigo	6	6	6	Yes					
Hiro	6	6	6	Yes					
Junko	6	6	6	Yes					
Hana	6	6	6	Yes					
Fuku	6	6	6	Yes					
Mama	6	6	6	Yes					
Toby	6	6	6	Yes					
Go	6	6	6	Yes					
Jin	6	6	6	Yes					
Tami	6	6	5	No	Rose	6	5	6	No
Ru	6	6	5	No	Jiro	6	6	4	No
Koara	6	6	5	No	Dan	6	4	6	No
Tomi	6	6	5	No	Miki	6	3	5	No
Riko	6	6	5	No	Kiko	6	4	5	No
Waka	6	6	5	No	Goro	6	5	5	No
Moto	6	6	5	No	Keiko	6	4	4	No
Ken	6	5	6	No	Yo	6	3	3	No
Mat	6	5	6	No					
Hide	6	5	5	No					

In the TG, for the remaining 10 students who did not maintain stage 6 on both posttests, the majority (9 of 10=90%) were assigned to stage 6 on one posttest and stage 5 on the other posttest, and only one was assigned to stage 5 on both posttests. These results suggest that the treatment was effective in helping a number of students: although they did not maintain stage 6 on both posttests, after receiving the treatment they began to produce more examples of productive stage 6 questions for two stage 6 Q-form. On the other hand, in the CG, of the eight students who did not maintain stage 6 on both posttests, although three students (37%) were assigned to stage 6 on one posttest and to either stage 4 or 5 on the other posttest, five students (63%) regressed by one or more stages on the posttests. This means that these five students did not produce two or more examples for two different stage 6 Q-forms on both posttests. For instance, one (Yo) regressed by three stages from stage 6 to stage 3 on both posttests, one (Keiko) regressed by two stages to stage 4 on both posttests, and one (Goro) regressed by one stage to stage 5. Given that none of the 22 stage 6 students in the TG was assigned to stage 4 or lower on the posttests, despite the use of the stricter criterion for the posttests, it is possible to conclude that the difference between the two conditions was brought about by the instruction.

It is of interest to point out that, if I had employed the less rigorous criterion used for the stage assessment for the pretest (i.e. two different questions for the given stage) in the stage assessment for the posttest, then 28 of the 31 students who had reached stage 6 on the pretest would have been considered to have maintained stage 6 on both posttests. This shows clearly how results in the stage gain analysis can be affected by how one defines *acquisition* (i.e. *emergence* of Q-forms). Both in this study and in Mackey and Philp (1998), in order to increase the probability that gain in developmental stage actually occurred, the stricter requirement (i.e. two productive examples of two different Q-forms) was employed for developmental stage assignment for the posttests. It is possible to say that the use of this more rigorous criterion served both to prevent misjudgement in the sustained stage gain assessment which can be caused by learners' use of formulaic structures, and to measure the effects of the treatment in a meaningful manner – in terms of L2 development. It should be noted that Mackey and Philp (1998) excluded the students who had reached stage 6 on the pretest in their analysis, so it is not possible to see what changes occurred in their production of questions after treatment.

I should point out that the results of the stage gain analysis do not show us whether the decrease in

developmental stage observed among the 16 students who had reached the highest stage on the pretest (TG=10, CG=8) were caused by their non-production of examples for two stage 6 rules, or by that for only one stage 6 rule. The same thing can be said for the seven Stage6-Ready students (TG=4, CG=3) who were considered not to have gained a stage, as discussed in the previous section. It is because the stage gain assessment was measured by individuals' *overall* developmental stage increase on the posttests, in relation to the pretest. The measurement of learners' changes in developmental stage is an effective way to assess the effects of the treatment, but there is a need for a detailed examination of individual's changes in producing examples for each target Q-form after the treatment, which will be taken up in the next chapter.

6.5 Summary

There were two research questions to examine the effects of the treatment on individuals' L2 development. The first was:

Q1. Is the *noticing –promotion approach* effective in helping the students advance in developmental stage in interrogative acquisition?

I have shown that the treatment was effective in promoting learners' advancement in developmental stage in the acquisition of question formation; in the TG five of the nine stage 6 Ready students (56%) increased a stage from 5 to 6 and sustained this gain over a period of some seven weeks. To the best of my knowledge, the longest sustained effect of instruction reported in previous studies was one month (Mackey & Philp, 1998), only about half the length of this study, which makes these learners' achievements particularly striking. In addition, another two stage 6 Ready advanced to stage 6 on the follow-up test, showing delayed effects of the treatment. On the other hand, such positive changes were not observed in the three stage6-Ready students who did not receive the treatment. These results do underscore the effectiveness of the approach used in this study for promoting L2 development in the acquisition of higher-level Q-forms, but the participant number in each condition was of course small and we should be cautious about drawing a firm conclusion.

Q2. Are there any differences in individuals' sustained stage gain in relation to their current developmental stage?

Whereas either the sustained effects of the treatment or the delayed effects were observed in the majority of stage 6-Ready students (7 of 9) who received the treatment, such positive change was seen in only two of the 16 stage 6-Unready students who received the same treatment. This suggests that the students' readiness for learning cognitively complex rules determined their acquisition of the target Q-forms, giving support to the Teachability Hypothesis. This reminds us of the importance of teachers' targeting rules at the next stage of learners' development in teaching English interrogatives.

Apart from the effects of the treatment on learners' stage gain, I demonstrated that the treatment was also beneficial to help students produce more stage 6 questions on the posttests and maintain stage 6 for seven weeks. As we have seen, more than half the students in the TG (12 of 22) maintained stage 6 on both posttests, while in the CG only one of the nine students did so. Moreover, in the CG group more than half the students (5 of 9) regressed by one or more stages on the posttests, while in the TG such a phenomenon was observed in only one student and the other 21 students were assigned to stage 6 on one of the two posttest. These results indicate that the treatment also had positive effects on the production of more stage 6 questions among the students who did not maintain stage 6 on both posttests.

To sum up, the positive effects found both in the treatment group support my claim that the treatment which provided the learners with explicit teaching of grammar and opportunities to pay attention to their own output and encouraged peer feedback through proactive noticing tasks is an effective way to promote L2 development in the acquisition of question formation.

In the next two chapters, I will explore in more detail the effects of the treatment on individual students' production of each target Q-form.

CHAPTER 7
Effects of the treatment on filling gaps in the acquisition of Q-forms

In the previous chapter, I noted that some students did not produce examples for all rules at the same stage even after reaching the highest-level, stage 6. Such gaps have been observed in other studies (e.g. Pienemann, Johnston, & Brindley's ESL study, 1988; Meisel, Clahsen, & Pienemann's GSL study, 1981), but little research has been done to examine what type of instruction is effective in filling the gaps. This is the second aim of my study. To this end, I formulated the following two research questions:

Q3. Do the students show developmental gaps in interrogative acquisition?
Q4. If so, is the instruction effective in filling such gaps?

In Section 7.1 I will explain the procedures for identifying gaps in an individual's output on the pretest. In Section 7.2 I will show how I measured the effects of the treatment on filling the gaps on the posttests. In Section 7.3 I will present my analysis of sustained effects of the treatment on the acquisition of each target Q-form. I will also discuss sustained effects of the treatment on the acquisition of untargeted Q-forms in which students had shown gaps on the pretest.

7.1 Procedures for identifying individuals' gaps on the pretest

In order to identify gaps in individual students' development of question formation on the pretest, I used the results of coding question examples obtained in the developmental stage assessment. In this analysis, I employed the following criterion.

> In order for students to be considered to have acquired rule *x*, they had to produce at least *two* productive usages for this rule.

In the following analysis, I have excluded question examples of stage 6 *Y/N-inversion* - one of the four target rules - since it occurred too infrequently in the two conversation tasks given in the tests to allow examination of its acquisition.

To illustrate the procedures for identifying gaps on the pretest with actual data, I will show how I carried out the implicational analysis for one student (Seigo).

Seigo

In the developmental stage assessment for the pretest, Seigo was considered to have reached stage 6. Learners at stage 6 have acquired the processing operations necessary to produce all Q-forms and, if there are not two or more productive question examples of any rules, they are considered to have shown gaps in L2 development. Using the quantitative data (see Table 7.1 below) obtained in the previous chapter, I examined whether he had produced at least two productive examples for each rule on the pretest, and produced an implicational scale (see Table 7.2 below).

TABLE 7.1 Number of productive examples of each Q-form (Seigo)

	1	*2*	*3*	*4wh*	*4yn*	*5wh*	*5Nwh*	*6C*	*6T*
Pretest	3 (2)	0 (0)	13 (13)	3 (2)	12 (4)	6 (4)	2 (2)	0 (0)	3 (3)

TABLE 7.2 Result of the implicational analysis (Seigo)

	1	*2*	*3*	*4wh*	*4yn*	*5wh*	*5Nwh*	*6C*	*6T*
Pretest	+	/	+	+	+	+	+	-	+

In the implicational table, a '+' marks the presence of at least two productive usages of rules. For instance, there were three productive examples of *Q-tag* produced with three different auxiliary verbs (see below) and he was considered to have acquired this rule.

In your picture, err aren't there . err one cat . uh. In your picture, there are not there are not cats on the table, are you uh are there? ↑
In your picture, err there are three trees in the garden, aren't there? ↑
In in your picture, umm . the uh its girl wear yellow shirt, err err don't . doesn't she? ↑

Seigo also produced at least two productive examples for the following six rules: stage 1 Q-form, stage 3 Q-form, two stage 4 Q-forms and two stage 5 forms, which is marked by a '+' in Table 7.2.

In the implicational table, a '-' indicates that there was zero or only one example of rules in the presence of two or more obligatory occasions to apply these rules. For example, in Seigo's data there was no example of successful application of *Cancel-Inversion* and he failed to apply this rule in the presence of two obligatory occasions provided by the examiner's prompt to apply this rule, as quoted below.

*Could you tell me are there mm . uh could you tell me are there are there two trees in the garden?
*Could you tell me how many trees in the garden?

It should be noted that, even if these two *Cancel-Inversion* questions had been produced voluntarily without the examiner's prompt, they would have been used as evidence for the student's non-acquisition of this rule. In the same way that a learner's application of a certain rule serves as evidence of their acquisition of that rule, its non- or misapplication on the obligatory occasions for applying it proves very valuable in assessing a learner's acquisition of Q-forms. However, if Seigo had produced one *Cancel-Inversion* question without the presence of two or more obligatory occasions to apply this rule, his data would have been excluded from the analysis of effects of the treatment on filling the gap for this rule, since it is not possible to examine his acquisition of this rule without sufficient evidence. Such cases are marked by a '?' in the implicational analysis.

In the implicational table, '/' marks that there were no occasions to apply rules. For instance, Seigo had no obligatory occasions for applying stage 2 Q-form and this is marked by a '/'. The absence of question examples for this rule was not evidence for his non-acquisition of this rule. Thus, the quantitative data (i.e. Table 7.1) was converted into qualitative data in the implicational analysis (i.e. Table 7.2).

7.2 Procedures for the analysis of sustained effects of the treatment

After identifying gaps in each student's output on the pretest, I examined whether they were filled after the treatment by producing an implicational scale for each student and for each posttest. The following criterion was used for this assessment.

> In order for students to be considered to have acquired rule *x*, they had to produce:
>
> at least *two* productive usages for this rule
> on the *two* posttests.

This criterion is arbitrary, but I took a slightly conservative position and, for students to be considered to have acquired rules, they were required to sustain gains for 6-7 weeks until the follow-up test. As was the case in the stage gain assessment in Chapter 6, this requirement was intended to increase the probability that gaps had been filled through the treatment. To illustrate the procedures for the assessment with actual data, I will cite results of my implicational analysis for Waka.

Waka

On the pretest, Waka was assigned to stage 6, but produced only one example for *Cancel-Inversion* and *Q-tag* at stage 6 (see Table 7.3 below), showing gaps for these rules (see Table 7.4 below). She also showed gaps for stage 5 *Negative-Do/Aux-2nd* and stage 4 *Wh-Inversion* on this test.

TABLE 7.3 Number of productive examples of each Q-form (Waka)

	1	*2*	*3*	*4wh*	*4yn*	*5wh*	*5Nwh*	*6C*	*6T*	Assigned stage
Pretest	5 (3)	5 (5)	12 (7)	3 (1)	7 (2)	5 (3)	0 (0)	3 (1)	1 (1)	Stage 6
Posttest	3 (1)	4 (4)	5 (4)	3 (1)	6 (4)	8 (3)	2 (2)	4 (3)	4 (2)	Stage 6
Follow-up test	4 (2)	1 (1)	8 (7)	1 (1)	10 (2)	7 (2)	2 (2)	2 (1)	3 (3)	Stage 6

TABLE 7.4 Result of the implicational analysis (Waka)

	1	*2*	*3*	*4wh*	*4yn*	*5wh*	*5Nwh*	*6C*	*6T*
Pretest	+	+	+	-	+	+	-	-	-
Posttest	+	+	+	-	+	+	+	+	+
Follow-up test	+	+	+	-	+	+	+	-	+

To examine whether these gaps had been filled after the treatment, I examined Waka's question examples produced on each posttest. For *Cancel-Inversion*, on the posttest Waka produced three productive examples of this rule and was considered to have acquired this rule (marked by a '+' in the implicational table). On the follow-up test, although she produced two *Cancel-Inversion* questions, the same auxiliary verb was used in both questions and one was invalidated as evidence. Waka was considered not to have acquired this rule on the follow-up test (marked by a '-'). Waka did not sustain the gain on the posttest until the follow-up test and I concluded that her gap for this rule had been left unfilled.

For *Q-tag*, on both posttests, Waka produced two or more productive examples of this rule and this is marked a '+' for each posttest in the implicational analysis. For instance, on the posttest she produced four tag questions and two of them were considered productive examples and on the follow-up test she produced three productive examples with different auxiliary verbs. Based I these observations, I concluded that her gap for *Q-tag* had been filled.

For Waka's gap for stage 5 *Negative-Do/Aux-2nd*, on both posttests she produced two productive examples and the gap was considered to be filled. As to her gap for stage 4 *Wh-Inversion*, I concluded that it had been left unfilled, since there was only one piece of positive evidence for this rule on each posttest.

For each student showing the gaps on the pretest, I carried out an implicational analysis of this type.

7.3 Findings and discussion

7.3.1 Results of the analysis of developmental gaps on the pretest

The results of my implicational analysis of the 59 students' acquisition of Q-forms on the pretest are presented in Table 7.6 shown on the next page. At the top of the table, I list 59 students' pseudonyms, and above I indicate their given developmental stage for this test. On the left of the table, I list nine Q-forms based on the six-stage sequence of the acquisition of question formation. As we can see, many students showed gaps for rules at stage 4 and above, except for *Affirmative-Aux/Do-2nd* at stage 5. Based on the results of the implicational analysis, I summarise the number of students showing gaps for each rule in relation to their developmental stage on the pretest in Table 7.5 below.

TABLE 7.5 Number of students showing gaps (pretest)

	Stage 4		Stage 5		Stage 6	
	Wh-inversion	*Y/N-inversion*	*Affirmative-Aux/Do-2nd*	*Negative-Aux/Do-2nd*	*Cancel-inversion*	*Q-tag*
Treatment group (N=47)						
Stage 6 (n=22)	15	1	0	7	11	9
Stage 5 (n= 9)	7	0	0	5	n/a	n/a
Stage 4 (n= 4)	4	3	n/a	n/a	n/a	n/a
Stage 3 (n=11)	n/a	n/a	n/a	n/a	n/a	n/a
Stage 1 (n= 1)	n/a	n/a	n/a	n/a	n/a	n/a
Subtotal	26	4	0	12	11	9
Comparison group (N=12)						
Stage 6 (n=9)	4	0	0	7	7	4
Stage 5 (n=3)	3	0	0	2	n/a	n/a
Subtotal	7	0	0	9	7	4
TOTAL	33/47 (70%)	4/47 (9%)	0/43 (0%)	21/43 (49%)	18/31 (58%)	13/31 (42%)

TABLE 7.6 Implicational table (pretest)

Stage	Name	*6T*	*6C*	*5Nwh*	*5wh*	*4yn*	*4wh*	*3*	*2*	*1*
6	Yo	+	-	-	+	+	-	+	+	+
	Dan	-	+	-	+	+	-	+	+	+
	Goro	+	-	+	+	+	+	+	+	+
	Jiro	-	-	+	+	+	+	+	+	/
	Taro	+	-	-	+	+	-	+	+	+
	Kiko	-	-	-	+	+	-	+	+	+
	Miki	-	-	-	+	+	+	+	/	+
	Rose	+	+	-	+	+	+	+	+	+
	Keiko	+	-	-	+	+	+	+	+	+
	Moto	-	+	-	+	+	-	+	+	+
	Mat	-	+	+	+	-	+	+	/	/
	Jin	-	-	+	+	+	-	+	+	+
	Hide	+	-	+	+	+	-	+	+	+
	Go	+	-	+	+	+	-	+	+	+
	Toby	+	-	+	+	+	-	+	+	+
	Waka	-	-	-	+	+	-	+	+	+
	Mama	+	+	-	+	+	+	+	/	+
	Fuku	-	+	-	+	+	-	+	+	+
	Riko	-	+	+	+	+	-	+	+	+
	Hana	-	+	+	+	+	+	+	+	/
	Tomi	-	-	+	+	+	-	+	+	+
	Koara	+	+	+	+	+	+	+	+	+
	Junko	+	-	+	+	+	-	+	+	+
	Hiro	+	-	-	+	+	-	+	+	+
	Seigo	+	-	+	+	+	+	+	/	+
	Ru	+	-	+	+	+	-	+	+	+
	Yuta	+	-	+	+	+	-	+	+	+
	Yan	+	+	+	+	+	-	+	/	+
	Ken	+	+	-	+	+	-	+	+	+
	Tami	+	+	-	+	+	+	+	/	+
	Saya	-	+	+	+	+	+	+	/	+
5	Sae	-	-	-	+	+	-	+	+	+
	Mi	-	-	+	+	+	-	+	+	+
	Seto	-	-	-	+	+	-	+	+	+
	Taka	-	-	+	+	+	-	+	+	+
	Shin	-	-	-	+	+	+	+	+	+
	Kaji	-	-	-	+	+	-	+	+	+
	Yuya	-	-	-	+	+	-	+	+	+
	Nae	-	-	+	+	+	-	+	+	+
	Tana	-	-	+	+	+	+	+	+	+
	Jo	-	-	-	+	+	-	+	+	+
	Nobu	-	-	+	+	+	-	+	+	+
	Toyo	-	-	-	+	+	-	+	+	+
4	Mo	-	-	-	-	+	-	+	+	+
	Kako	-	-	-	-	-	-	+	/	+
	Rika	-	-	-	-	-	-	+	+	+
	Omi	-	-	-	-	-	-	+	+	+
3	Nori	-	-	-	-	-	-	+	+	+
	Tomo	-	-	-	-	-	-	+	+	+
	Ai	-	-	-	-	-	-	+	+	+
	Saki	-	-	-	-	-	-	+	+	+
	Sato	-	-	-	-	-	-	+	+	+
	Chie	-	-	-	-	-	-	+	+	+
	Maya	-	-	-	-	-	-	+	+	+
	Yaya	-	-	-	-	-	-	+	/	+
	Sue	-	-	-	-	-	-	+	+	+
	Yuko	-	-	-	-	-	-	+	+	+
	Eri	-	-	-	-	-	-	+	+	+
1	Tsuta	-	-	-	-	-	-	-	-	+

Legend: + acquired (two or more productive examples) - not acquired (one or no productive example) / no occasion to apply rule x

For two stage 6 Q-forms, on the pretest the 31 students (TG=22, CG=9) assigned to stage 6 were supposed to have acquired the required processing procedures to apply all rules at stage 6, but 18 of them (TG=11, CG=7; 58%) showed a gap for *Cancel-inversion*, and 13 of the 31 students (TG=9, CG=4; 42%) showed a gap for *Q-tag*. Of these 18 students, six showed gaps for both rules (Tomi, Waka & Jin in the TG; Miki, Kiko & Jiro in the CG). It should be noted that the remaining 28 students (TG=25, CG=3) assigned to stage 5 or lower on the pretest had not acquired the processing prerequisite for stage 6, so their non-acquisition of stage 6 forms was not considered to be a gap in their L2 development.

For two stage 5 Q-forms, on the pretest none of the 43 students at stage 5 or stage 6 (TG=31, CG=12) showed a gap for *Affirmative-Aux/Do-2nd*. In contrast, for the other stage 5 rule, *Negative-Aux/Do-2nd*, more than half (21 of 43=49%) showed a gap: 14 stage 6 students (TG=7, CG=7) and seven stage 5 students (TG=5, CG=2). The difference in the acquisition of two stage 5 rules suggests that the learners tended not to develop the two forms at the same time in their L2 development, although the same degree of processing complexity (i.e. the placement of an auxiliary in the second position to the *wh*/*Q*-word before the subject) is required to apply these two rules. As I noted earlier, two stage 5 Q-forms have been conflated as one rule in the previous studies, and this study is the first to report such a difference in the acquisition of the two rules.

For the two stage 4 Q-forms, of the 47 students at stage 4 or above on the pretest (TG=35, CG=12), four students (9%) showed a gap for *Y/N-inversion*. For *Wh-inversion*, more than two thirds of the students (33 of 47=70%; TG=26, CG=7) showed a gap and their developmental stage varied from stage 4 to stage 6 (4 stage 4 students, 10 stage 5students, 19 stage 6 students). This shows that the gap for *Wh-inversion* can be left unfilled, even after reaching the highest level in the acquisition of question formation. The treatment in this study focused on stage 6 rules and one stage 5 rule, and did not target these stage 4 forms; it is interesting to see whether or not the gaps for the untargeted rules can be filled through the treatment. In the next section, I will present my analysis of the sustained effect of the treatment on filling the gaps for the targeted Q-forms, and then I will compare the results with those obtained from the analysis of sustained effects on filling the gaps for the aforementioned two untargeted stage 4 rules.

7.3.2 Sustained effects of the treatment – the target Q-forms

In this section, I will examine effects of the treatment on filling the gaps for three target Q-forms separately (*Cancel-inversion, Q-tag , Negative-Aux/Do-2*[nd]).

7.3.2.1 *Cancel-inversion* (stage 6)

On the pretest, there were 31 students (TG=22, CG=9) assigned to stage 6 in two conditions. They were supposed to have acquired the processing prerequisite for stage 6, but gaps for stage 6 *Cancel-inversion* (6C) were observed in 18 students (TG=11, CG=7), which is marked by '-' in their implicational scales shown in Table 7.7 below.

TABLE 7.7 Implicational scales for the students showing a gap for *Cancel-inversion*

[Treatment group]

Stage 6		*1*	*2*	*3*	*4wh*	*4yn*	*5wh*	*5Nwh*	*6C*	*6T*	Gap filled?
Yuta	Pre	+	+	+	-	+	+	+	-	+	
Yuta	Post	+	+	+	-	+	+	+	+	+	YES
Yuta	Follow-up	+	+	+	?	+	+	+	+	+	
Ru	Pre	+	+	+	-	+	+	+	-	+	
Ru	Post	+	+	+	+	+	+	+	+	+	YES
Ru	Follow-up	+	+	+	-	+	+	+	+	-	
Seigo	Pre	+	/	+	+	+	+	+	-	+	
Seigo	Post	+	+	+	+	+	+	+	+	+	YES
Seigo	Follow-up	+	+	+	+	+	+	+	+	+	
Hiro	Pre	+	+	+	-	+	+	-	-	+	
Hiro	Post	+	/	+	+	+	+	-	+	+	YES
Hiro	Follow-up	+	+	+	+	+	+	+	+	+	
Junko	Pre	+	+	+	-	+	+	+	-	+	
Junko	Post	+	+	+	+	+	+	-	+	+	YES
Junko	Follow-up	+	+	+	+	+	+	-	+	+	
Tomi	Pre	+	+	+	-	+	+	+	-	-	
Tomi	Post	+	+	+	-	+	+	+	+	+	YES
Tomi	Follow-up	+	/	+	+	+	+	+	+	-	
Waka	Pre	+	+	+	-	+	+	-	-	-	
Waka	Post	+	+	+	-	+	+	+	+	+	
Waka	Follow-up	+	+	+	-	+	+	+	-	+	
Toby	Pre	+	+	+	-	+	+	+	-	+	
Toby	Post	+	+	+	-	+	+	+	+	+	YES
Toby	Follow-up	+	+	+	+	+	+	+	+	+	
Go	Pre	+	+	+	-	+	+	+	-	+	
Go	Post	+	+	+	-	+	+	-	+	+	YES
Go	Follow-up	/	+	+	?	+	+	-	+	+	
Hide	Pre	+	+	+	-	+	+	+	-	+	
Hide	Post	/	+	+	-	+	+	+	-	+	
Hide	Follow-up	+	+	+	+	+	+	+	-	+	
Jin	Pre	+	+	+	-	+	+	+	-	-	
Jin	Post	+	+	+	-	+	+	-	+	+	YES
Jin	Follow-up	+	+	+	?	+	+	+	+	+	

[Comparison group]

Stage 6		*1*	*2*	*3*	*4wh*	*4yn*	*5wh*	*5Nwh*	*6C*	*6T*	Gap filled?
Keiko	Pre	+	+	+	+	+	+	-	-	+	
Keiko	Post	+	+	+	+	+	+	-	-	+	
Keiko	Follow-up	+	+	+	+	+	+	-	-	-	
Miki	Pre	+	/	+	+	+	+	-	-	-	
Miki	Post	+	/	+	-	+	+	-	-	-	
Miki	Follow-up	+	+	+	+	+	+	+	-	-	
Kiko	Pre	+	+	+	-	+	+	-	-	-	
Kiko	Post	+	+	+	+	+	+	-	-	-	
Kiko	Follow-up	+	+	+	?	+	+	+	-	+	
Taro	Pre	+	+	+	-	+	+	-	-	+	
Taro	Post	+	+	+	+	+	+	+	+	+	YES
Taro	Follow-up	+	+	+	+	+	+	-	+	+	
Jiro	Pre	/	+	+	+	+	+	+	-	-	
Jiro	Post	+	+	+	-	+	+	-	+	+	
Jiro	Follow-up	+	+	+	+	+	+	-	-	+	
Goro	Pre	+	+	+	+	+	+	+	-	+	
Goro	Post	+	+	+	+	+	+	+	-	-	
Goro	Follow-up	+	+	+	+	+	+	+	-	+	
Yo	Pre	+	+	+	-	+	+	-	-	-	
Yo	Post	+	+	+	-	+	+	-	-	+	
Yo	Follow-up	+	+	+	-	+	+	-	-	-	

On the pretest, most examples of *Cancel-inversion* were produced at the end of the Spot-the-Difference task in response to the examiner's prompt to use this rule, while 12 students voluntarily produced *Cancel-inversion* questions without any prompts. In the TG, nine of the 11 students gained a '+' for this rule on the two posttests (Yuta, Ru, Seigo, Hiro, Junko, Tomi, Toby, Go, Jin) by producing two or more productive examples of this rule, while one (Hide) got a '-' again for both posttests and one student (Waka) gained a '+' only on the posttest.

In contrast, in the CG, of the seven stage 6 students showing gaps for *Cancel-inversion* on the pretest, only one (Taro) gained a '+' on both posttests. The rest of five students (Keiko, Miki, Kiko, Goro, Yo) got a '-' on both posttests and one student (Jiro) got a '+' only on the immediate posttest.

I summarise these results in Table 7.8 below. The majority of TG students' (9 of 9=82%) gaps were filled, while the majority of CG students' (1 of 7=14%) gaps were left unfilled. The difference between the two conditions shows that the treatment was effective in helping students fill the gaps for *Cancel-inversion*.

TABLE 7.8 Effects of the treatment on filling the gaps for *Cancel-inversion* (Stage 6)

		Number of students who showed gaps on the pretest	Number of students whose gaps were filled	
Treatment group	Stage 6 (N=22)	11	9	(82%)
Comparison group	Stage 6 (N=9)	7	1	(14%)

7.3.2.2 *Q-tag* (stage 6)

For stage 6 *Q-tag*, similar results were observed. On the pretest most examples of *Q-tag* were produced at the end of the Spot-the-Difference task by the examiner's prompt to use this rule, although five students voluntarily produced tag questions without any prompts. In the TG, on the two post tests, seven of the nine students showing a gap for this rule (Saya, Hana, Fuku, Waka, Jin, Mat, Moto) produced two or more productive examples of *Q-tag*, while two students (Tomi, Riko) got a '+' only on

one of the post tests, as shown in Table 7.9 below.

On the other hand, in the CG, only one (Jiro) of the four students showing a gap for this rule on the pretest gained a '+' for both posttests, while one (Miki) got a '-' on both tests and two (Kiko, Dan) got a '+' only on the follow-up test.

TABLE 7.9 Implicational scales for the students showing a gap for *Q-tag*

[Treatment group]

Stage 6		1	2	3	4wh	4yn	5wh	5Nwh	6C	6T	Gap filled
Saya	Pre	+	/	+	+	+	+	+	+	-	
Saya	Post	+	+	+	+	+	+	+	+	+	YES
Saya	Follow-up	+	+	+	-	+	+	+	+	+	
Tomi	Pre	+	+	+	-	+	+	+	-	-	
Tomi	Post	+	+	+	-	+	+	+	+	+	
Tomi	Follow-up	+	/	+	+	+	+	+	+	-	
Hana	Pre	/	+	+	+	+	+	+	+	-	
Hana	Post	/	+	+	-	+	+	+	+	+	YES
Hana	Follow-up	/	+	+	-	+	+	+	+	+	
Riko	Pre	+	+	+	-	+	+	+	+	-	
Riko	Post	+	+	+	-	+	+	+	-	-	
Riko	Follow-up	+	+	+	-	+	+	+	-	+	
Fuku	Pre	+	+	+	-	+	+	-	+	-	
Fuku	Post	+	+	+	-	+	+	-	+	+	YES
Fuku	Follow-up	+	+	+	+	+	+	-	+	+	
Waka	Pre	+	+	+	-	+	+	-	-	-	
Waka	Post	+	+	+	-	+	+	+	+	+	YES
Waka	Follow-up	+	+	?	-	+	+	+	-	+	
Jin	Pre	+	+	+	-	+	+	+	-	-	
Jin	Post	+	+	+	-	+	+	-	+	+	YES
Jin	Follow-up	+	+	+	-	+	+	+	+	+	
Mat	Pre	/	/	+	+	-	+	+	+	-	
Mat	Post	/	+	+	+	+	+	+	-	+	YES
Mat	Follow-up	/	+	+	+	+	+	+	+	+	
Moto	Pre	+	+	+	-	+	+	-	+	-	
Moto	Post	+	+	+	+	+	+	-	+	+	YES
Moto	Follow-up	+	+	+	+	+	+	+	-	+	

[Comparison group]

Stage 6		1	2	3	4wh	4yn	5wh	5Nwh	6C	6T	Gap filled
Miki	Pre	+	/	+	+	+	+	-	-	-	
Miki	Post	+	/	+	-	+	+	-	-	-	
Miki	Follow-up	+	+	+	+	+	+	+	-	-	
Kiko	Pre	+	+	+	-	+	+	-	-	-	
Kiko	Post	+	+	+	+	+	+	-	-	-	
Kiko	Follow-up	+	+	+	?	+	+	+	-	+	
Jiro	Pre	/	+	+	+	+	+	+	-	-	
Jiro	Post	+	+	+	-	+	+	-	+	+	YES
Jiro	Follow-up	+	+	+	+	+	+	-	-	+	
Dan	Pre	+	+	+	-	+	+	-	+	-	
Dan	Post	+	+	+	+	+	+	-	-	-	
Dan	Follow-up	+	+	+	-	+	+	-	+	+	

These results are summarised in Table 7.10 on the next page. In the TG seven of the nine students' gaps were filled (78%), while in the CG only one of the four students' gaps were filled (25%). The larger proportion observed in the TG suggests that the treatment was beneficial to help learners who had reached stage 6 acquire *Q-tag*.

TABLE 7.10 Effects of the treatment to fill the gaps for *Q-tag* (Stage 6)

		Number of students who showed gaps on the pretest	Number of students whose gaps were filled	
Treatment group	Stage 6 (N=22)	9	7	(78%)
Comparison group	Stage 6 (N=9)	4	1	(25%)

7.3.2.3 *Negative-Aux-2nd* (Stage 5)

I now turn to the effects of the treatment to fill the gap for stage 5 *Negative-Aux/Do-2nd*. As predicted, on the pretest all examples of this rule were produced on the embedded occasions for applying this rule, and no students voluntarily created occasions for applying it on their own. This underlines the difficulty of eliciting question examples of *Negative-Aux/Do-2nd* without seeding occasions for applying this rule.

As mentioned earlier, the 43 students assigned to stage 5 or above on the pretest (TG=31, CG=12) were supposed to have acquired the processing prerequisite for stage 5, but 21 students (TG=12, CG=9) showed a gap for *Negative-Aux/Do-2nd*. In the following analysis I excluded two students' data (Moto in the TG and Taro in the CG) because application of the criterion, which required the presence of two productive examples as evidence for the acquisition of Q-forms, was not feasible due to the lack of obligatory occasions for applying this rule in their samples. They produced two productive examples of this rule on one of the two posttests and would have been considered to have acquired the rule, had they produced two productive examples of the rule on the other test; however, they had only one obligatory occasion on the other test in which they applied this rule appropriately. After the exclusion of the two students, the number of students showing a gap for this rule in each condition was 11 students in the TG and eight students in the CG (see Table 7.11 on the next page).

TABLE 7.11 Implicational scales for the students showing a gap for *Negative-Aux/Do-2nd*

[Treatment group]

Stage 6		1	2	3	4wh	4yn	5wh	5Nwh	6C	6T	Gap filled
Tami	Pre	+	/	+	+	+	+	-	+	+	
Tami	Post	+	+	+	+	+	+	+	+	+	YES
Tami	Follow-up	/	+	+	+	+	+	+	-	+	
Ken	Pre	+	+	+	-	+	+	-	+	+	
Ken	Post	/	+	+	+	+	+	+	-	+	YES
Ken	Follow-up	+	+	+	-	+	+	+	+	+	
Hiro	Pre	+	+	+	-	+	+	-	-	+	
Hiro	Post	+	/	+	+	+	+	-	+	+	
Hiro	Follow-up	+	+	+	+	+	+	+	+	+	
Fuku	Pre	+	+	+	-	+	+	-	+	-	
Fuku	Post	+	+	+	-	+	+	-	+	+	
Fuku	Follow-up	+	+	+	+	+	+	-	+	+	
Mama	Pre	+	/	+	+	+	+	-	+	+	
Mama	Post	+	+	+	+	+	+	-	+	+	
Mama	Follow-up	+	/	+	+	+	+	-	+	+	
Waka	Pre	+	+	+	-	+	+	-	-	-	
Waka	Post	+	+	+	-	+	+	+	+	+	YES
Waka	Follow-up	+	+	+	-	+	+	+	-	+	
Stage 5											
Toyo	Pre	+	+	+	-	+	+	-	-	-	
Toyo	Post	+	+	+	?	+	+	-	-	+	
Toyo	Follow-up	+	+	+	-	+	+	-	+	+	
Jo	Pre	+	+	+	-	+	+	-	-	-	
Jo	Post	+	+	+	-	+	+	-	+	+	
Jo	Follow-up	+	+	+	-	+	+	-	+	+	
Yuya	Pre	+	+	+	-	+	+	-	-	-	
Yuya	Post	/	+	+	+	+	+	+	-	+	YES
Yuya	Follow-up	+	/	+	-	+	+	+	+	+	
Kaji	Pre	+	+	+	-	+	+	-	-	-	
Kaji	Post	+	+	+	+	+	+	+	+	+	YES
Kaji	Follow-up	/	+	+	?	+	+	+	+	+	
Shin	Pre	+	+	+	+	+	+	-	-	-	
Shin	Post	+	+	+	-	+	+	-	-	+	
Shin	Follow-up	+	+	+	-	-	+	/	+	-	

[Comparison group]

Stage 6		1	2	3	4wh	4yn	5wh	5Nwh	6C	6T	Gap filled
Keiko	Pre	+	+	+	+	+	+	-	-	+	
Keiko	Post	+	+	+	+	+	+	-	-	+	
Keiko	Follow-up	+	+	+	+	+	+	-	-	-	
Rose	Pre	+	+	+	+	+	+	-	+	+	
Rose	Post	+	+	+	+	+	+	+	-	+	YES
Rose	Follow-up	+	+	+	+	+	+	+	+	+	
Miki	Pre	+	/	+	+	+	+	-	-	-	
Miki	Post	+	/	+	-	+	+	-	-	-	
Miki	Follow-up	+	+	+	+	+	+	+	-	-	
Kiko	Pre	+	+	+	-	+	+	-	-	-	
Kiko	Post	+	+	+	+	+	+	-	-	-	
Kiko	Follow-up	+	+	+	?	+	+	+	-	+	
Dan	Pre	+	+	+	-	+	+	-	+	-	
Dan	Post	+	+	+	+	+	+	-	-	-	
Dan	Follow-up	+	+	+	-	+	+	-	+	+	
Yo	Pre	+	+	+	-	+	+	-	-	-	
Yo	Post	+	+	+	-	+	+	-	-	+	
Yo	Follow-up	+	+	+	-	+	+	-	-	-	
Stage 5											
Seto	Pre	+	+	+	-	+	+	-	-	-	
Seto	Post	+	+	+	+	+	+	-	-	-	
Seto	Follow-up	+	+	+	+	+	+	-	-	+	
Sae	Pre	+	+	+	-	+	+	-	-	-	
Sae	Post	+	+	+	-	+	+	-	-	+	
Sae	Follow-up	+	+	+	-	+	+	-	-	+	

In the TG, on both posttests five of the 11 students showing a gap for *Negative-Aux/Do-2nd* (Yuya and Kaji at stage 5 and Tami, Ken, Waka at stage 6) gained a '+', while four students (Toyo and Jo at stage 5 and Fuku and Mama at stage 6) got a '-' on both tests and one student (Shin at stage 5) got a '-'on the posttest and got a '/' on the follow-up test. One student (Hiro at stage 6) gained a '+' only on the follow-up test.

On the other hand, in the CG only one of the eight students (Rose at stage 6) got a '+' on both posttests, while five students (Seto and Sae at stage 5 and Keiko, Dan and Yo at stage 6) got a '-' for both posttests and two students (Miki and Kiko at stage 6) gained a '+' only on the follow-up test. I summarise these

results in Table 7.12 below.

TABLE 7.12 Effects of the treatment to fill the gaps for *Negative-Aux/Do-2nd* (Stage 5)

	Developmental stage on the pretest	Number of students who showed gaps on the pretest	Number of students whose gaps were filled
Treatment group	Stage 6 (N=22)	6	3
	Stage 5 (N=9)	5	2
	Total	11	5 (45%)
Comparison group	Stage 6 (N=9)	6	1
	Stage 5 (N=3)	2	0
	Total	8	1 (13%)

In the TG, five of the 11 student's (45%) gaps were filled, while in the CG only one of the eight students' (13%) gap was filled. The difference between the two conditions suggests that the treatment was effective in helping the students fill the gaps for *Negative-Aux/Do-2nd*, but its effect was smaller than that observed for the other two target rules at stage 6 (*Cancel-inversion*: 9 of 11=82%, *Q-tag*: 6 of 8=75%). Three reasons can be suggested for this. Firstly, the smaller effect of the treatment might be caused by fewer opportunities to use *Negative-Aux/Do-2nd* than the other target forms in the treatment. *Negative-Aux/Do-2nd* was taught in the last session (Week 4) together with stage *6 Y/N-negative* structure, while *Cancel-inversion* and *Q-tag* were taught separately in the first two sessions. In addition, they had opportunities to use these two stage 6 rules in the third session and the amount of time spent on learning the stage 6 forms was longer than that on learning stage 5 *Negative-Aux/Do-2nd*.

Secondly, there was a possibility that, as discussed in Section 3.3.3.1, many students had been on a falling curve of the learning U-curve in the acquisition of *Negative-Aux/Do-2nd* when they took the posttest, which was given one week after the last treatment session in which this rule was taught. In addition, in the TG, delayed acquisition of this rule was observed in only one of the six students (Hiro). For two stage 6 rules (*Cancel-inversion* and *Q-tag*), they were taught in the first three sessions and the students had more time to take them up in their L2 development. If *Negative-Aux/Do-2nd* had been

taught in an earlier session of the treatment, more students might have performed better in using this form on the posttest, as they did in the acquisition of two stage 6 rules.

Third, it is also conceivable that the inclusion of two Q-forms at different levels in the last treatment session (stage 5 *Negative-Aux/Do-2nd* and stage 6 *Y/N-negative*) had a negative effect on learning. I taught the two negative Q-forms by contrasting the difference in their usage, since I thought it would be an effective way to help students understand appropriate contexts for using each form. However, about half the 11 students' (6 students) gap for *Negative-Aux/Do-2nd* was not filled through the treatment. It should also be noted that by including two target Q-forms the amount of time spent on practising each Q-form in this session was less than half that spent on the other target rules and this might have also had a negative influence on the acquisition of *Negative-Aux/Do-2nd*. I do not have enough data to examine the acquisition of the other target Q-form (i.e. *Y/N-negative*), and it is not possible to determine whether the smaller effects of the treatment on filling gaps in the acquisition of *Negative-Aux/Do-2nd* were caused by teaching two Q-forms at different stages in one class, or by the shorter amount of time spent by the learners practising this rule. It seems reasonable to speculate that the smaller effects of the treatment were caused by both elements.

To sum up, it is not possible to determine whether any of these reasons had more negative effects than the others on the acquisition of *Negative-Aux/Do-2nd*, but it seems reasonable to argue that the combination of these three causes led to the smaller effects of the treatment. The fact that more students showed a gap for *Negative-Aux/Do-2nd* than the other two target rules, and that the treatment was more effective in filling the gaps observed for the other two target Q-forms, suggests a need for more studies to examine the acquisition of *Negative-Aux/Do-2nd* and effects of instruction to teach this rule.

7.3.2.4 Acquisition of two stage 5 Q-forms

In addition to the lowest rate observed in the effects of the treatment on the acquisition of stage 5 *Negative-Aux/Do-2nd* among three target rules, I should point out the difference observed in the acquisition of two stage 5 Q-forms: *Affirmative-* and *Negative-Aux/Do-2nd*. These two rules require the

same degree of processing complexity (i.e. placement of an auxiliary in the second position of the sentence before the subject) and learners who have reached stage 5 are considered able to produce examples of both Q-forms. Contrary to this, we have seen that none of the 43 students at stage 5 or stage 6 on the pretest showed a gap for *Affirmative-Aux/Do-2nd* on the pretest, while about half of the 43 students (21 students) showed a gap for its negated form, *Negative-Aux/Do-2nd*. One reason suggests itself. L2 learners have more occasions for using *Affirmative-Aux/Do-2nd* than *Negative-Aux/Do-2nd* inside and outside the classroom, and their unfamiliarity with applying *Negative-Aux/Do-2nd* might have added extra cognitive load to learners' access to this Q-form in the embedded contexts for applying this rule. In that case, what kind of additional cognitive work was required to apply the unfamiliar rule? One student's comment made at the end of the treatment session in which I taught this rule, may provide some insight.

Yuya, in one of the university treatment groups, showed a gap for this rule on the pretest. He told me that he had not used *Negative-Aux/Do-2nd* before the treatment, although he knew the rule. In the conversation task in pairs in the treatment, he tried out the rule and noticed that he managed to use it correctly as in 'Why haven't you finished ironing yet?', but it sounded 'strange' to him and he struggled to convince himself that what he had said was correct (i.e. his utterance including the inversion of the subject and the negated form of auxiliary). By 'strange' Yuya meant that the sentence 'Why you haven't finished ironing yet?' in which the subject and the negated form of auxiliary were not inverted would sound more natural to him. I cite the examples in which Yuya tried to apply *Negative-Aux/Do-2nd* on the three tests below. On the pretest, he tried to apply this rule three times and failed in his first two attempts, where he did not invert the subject and the negated form of auxiliary. These observations 'chime with' his comments on his lack of confidence in usage.

Pretest

*Why err why err . he man err err . can't err eat dinner at home?
*Why err the man haven't cleaned the bathroom?
Why why can't you . Why can't I why can't I use the washing machine at the moment?

After the treatment, Yuya had five occasions for applying this rule on the posttest and produced four productive usages of the rule, although he failed to realise the inversion in one instance. On the follow-up test, he had three occasions for applying this rule and used it correctly every time. Moreover,

we can see that Yuya made self-corrections to his mistake in realising inversion in the last example produced on each posttest. The increased positive evidence for the rule application of *Negative-Aux/Do-2nd* on both the posttests suggests that acquisition of the rule had occurred through the treatment.

Posttest

Why can't we use the microwave oven?
Why why can't I use the 24-hour shop um at the night?
Why have you . why have. Why haven't you send . a /application/ form for the scholarship to the university?
*Why the boy can't eat cake?
Why err why the man err sitting on the sofa err . Why err haven't the man sitting on the . sofa bought the skis?

Follow-up test

Why haven't you send a err . send that?
Why does the uh. Why doesn't the girl uh *chigau [=no]*. Why can't the girl go camping tomorrow?
Why the why the woman err. Why haven't the woman seen Tom?

As discussed in Chapter 6, in previous studies two stage 5 rules were conflated as one Q-form, and no studies have reported such a difference in the acquisition of *Negative-Aux/Do-2nd* and its affirmative form. It is not clear from the previous studies (e.g. Pienemann, Johnston, & Brindley, 1988; Mackey, 1995; Mackey & Philp, 1998) how many examples of *Negative-Aux/Do-2nd* were found in their corpora, but given that question examples were collected either in interviews or in the information-gap tasks without seeding occasions for applying this rule, the occurrence of *Negative-Aux/Do-2nd* questions was arguably very small. The present study has revealed the problem that many learners can develop the gap for *Negative-Aux/Do-2nd*; this is an intriguing finding and more studies should be undertaken to examine how this rule is acquired.

Thus, this study has revealed a difference in the acquisition of two stage 5 forms; some may argue that the difference was caused by the way examples of *Negative-Aux/Do-2nd* were elicited and that the results of this study are not informative. Indeed, all examples of *Negative-Aux/Do-2nd* were produced on the seeded occasions for applying this rule, while examples of *Affirmative-Aux/Do-2nd* were produced both on the seeded and unseeded occasions, which might have influenced the results. In the Spot-the-Difference task, for instance, the students read the incomplete sentence in the speech balloons,

such as 'I haven't bought skis because [ ]', and were required to elicit the missing information by asking questions. In such a context, being provided with the written incomplete sentence, students' attention may have been drawn more to *form* than to *meaning*, because they were looking at the given words in the sentence while producing questions. On the other hand, when producing *Affirmative-Aux/Do-2nd* questions like 'What color of earring is the lady wearing?' in the same task, their use of the rule was not conditioned by any written prompt, so their attention was drawn more to *meaning* than to *form*, although there was a possibility that the learners switched their attention to *form* from time to time, when the need arose while producing questions even in such a context. I should also point out that the learners had more chances to apply *Negative-Aux/Do-2nd* correctly if they paid attention to *form*, especially when provided with necessary words for forming questions, as was the case on the seeded occasions in the Spot-the-Difference task. The way *Negative-Aux/Do-2nd* was elicited might have had some effect on students' production of examples of this rule, but the argument that this was the reason for the difference in the acquisition of two stage 5 rules seems unconvincing. There is a need for further research to examine the acquisition of these stage 5 forms and to find ways to elicit examples for *Negative-Aux/Do-2nd* without, as far as possible, drawing learners' attention to *form*.

A careful study of the question examples collected in the two conversation tasks designed to elicit examples of this rule, suggests two ideas for improving the elicitation methods used in these tasks. Before introducing those ideas, I will briefly describe how I seeded these occasions to prompt the application of *Negative-Aux/Do-2nd* in the Guided Role-play. On the pretest, the students were required to ask if there was a washing machine in the flat in Question 15 (see Appendix 70) and heard the reply, 'Yes, there is, but we can't use it at the moment'. This reply was meant to prompt students to apply *Negative-Aux/Do-2nd* as in 'Why can't we use it?' in the following Question 16; however, contrary to my expectation, many students simply asked 'Why?'. After the treatment, students began to use *Negative-Aux/Do-2nd* on such seeded occasions more often than before, although many still simply asked 'Why?'. On the posttests, some also applied *Cancel-inversion* on these occasions as in 'Could you tell me why we can't use the guestroom?'. This was because the ultimate choice of Q-form to apply on those occasions was up to the students, since they were not instructed to use specific Q-forms in those contexts. I should point out that students' non-application of *Negative-Aux/Do-2nd* on those occasions was not used as negative evidence in the assessment of stage assignment and of the acquisition

of this rule, since native speakers can, of course, also ask just 'Why?' or use *Cancel-inversion* on such occasions.

When I piloted the testing tasks with the students from the General English Course at the University of Edinburgh, most applied *Negative-Aux/Do-2nd* on the seeded occasions in both tasks, and I did not anticipate that many students in the main study would not apply this rule in the Guided Role-play. If I had noticed the problem then, I would have made changes to my task design. My idea for enhancing the elicitation method used in the Guided Role-play is to seed an occasion for asking a reason why they were not able to use the washing machine - not immediately after hearing the reply for Q15, but later at the end of the task. By inserting a few occasions for asking questions not related to the reply for Q15, it is possible to create an occasion requiring students to ask a question in a full sentence. For instance, if I had seeded an occasion after the last guided question of the task (i.e. Q21: Ask when you can see Mary on Friday) and given an instruction that 'You've realised that you forgot to ask why the residents in the flat cannot use the washing machine at the moment. Ask a reason by starting your question with 'Oh, Kate, I forgot to ask something. Why [ ]?'. The provision of the *wh*-word 'why' at the beginning of a question should make use of *Negative-Aux/Do-2nd* obligatory. The placement of a *Fronting* element in the sentence-initial position is stage 3 rule and the provision of the *wh*-word 'why' will not harm the elicitation of stage 5 questions.

My solution for improving the elicitation method used in the Spot-the-Difference task was the use of the students' L1 (Japanese) for the texts in the speech balloons on the task sheet. At the beginning of this task, the learners were required to elicit a missing piece of information in the five speech balloons by asking a question. For instance, on the pretest the students read the incomplete English sentence 'I can't eat dinner at home because [ ].' in one of the speech balloons intended to prompt the learners to apply *Negative-Aux/Do-2nd* as in 'Why can't the boy eat dinner at home?'. If I had used Japanese in this speech balloon, I could have created an occasion which would have required students to ask a question in a full sentence and could have avoided the problem caused by the presentation of the English words. In the ESL context, the use of the learners' L1 in the test instruction is avoided and English is used throughout the task, but in the EFL context in which learners share an L1, such use of the L1 can help teachers and researchers to devise elicitation methods for specific language items.

Next, I will compare the effects of the treatment observed in the acquisition of three target Q-forms with those observed in the acquisition of two untargeted stage 4 Q-forms.

7.3 3 Effect of the treatment – untargeted Q-forms

On the pretest, 47 of the 59 students were assigned to stage 4 or above and they were supposed to have acquired the processing prerequisite for stage 4. However, gaps were found for both stage 4 Q-forms; involving four students (9%) for *Y/N-inversion* and 33 students (70%) for *Wh-inversion*, as discussed in Section 7.3.1. These results show that these two rules had not been acquired at the same time by many students. I will now examine the effectiveness of the treatment on filling the gap for each untargeted rule.

7.3.3.1 *Y/N-inversion* (stage 4)

The four students showing a gap for *Y/N-inversion* comprised three stage 4 students and one stage 6 student - all in the treatment group (see Table 7.13 below).

TABLE 7.13 Implicational scales for the students showing a gap for *Y/N-inversion*

[Treatment group] [Comparison group]

Stage 4		1	2	3	4wh	4yn	5wh	5Nwh	6C	6T	Gap filled
Omi	Pre	+	+	+	-	-	-	-	-	-	
Omi	Post	+	+	+	-	+	-	-	-	-	
Omi	Follow-up	+	+	+	-	-	-	-	-	+	
Rika	Pre	+	+	+	-	-	-	-	-	-	
Rika	Post	+	+	+	+	+	-	-	-	-	
Rika	Follow-up	+	+	+	-	-	-	-	-	-	
Kako	Pre	+	/	+	-	-	-	-	-	-	
Kako	Post	/	+	+	-	-	-	-	+	-	
Kako	Follow-up	+	+	+	-	-	-	-	-	-	
Stage 6											
Mat	Pre	/	/	+	+	-	+	+	+	-	
Mat	Post	/	+	+	+	+	+	+	-	+	YES
Mat	Follow-up	/	+	+	+	+	+	+	+	+	

On both posttests, one stage 6 student (Mat) produced two productive examples for this rule (marked by a '+' for each posttest) and his gap for *Y/N-inversion* was filled, but the gap observed in the three students at stage 4 was left unfilled. For instance, one student (Kako) got a '-' for both posttests and two students (Omi, Rika) gained a '+' only for the immediate posttest. I summarise these results in Table 7.14 below. The proportion of students considered to have acquired *Y/N-inversion* through the

treatment was 25 percent (1 of 4), smaller than that found in the effects of the treatment to acquire the target forms (*Negative-Aux/Do-2nd*: 5 of 11=45%, *Cancel-inversion*: 9 of 11=82%, *Q-tag*: 6 of 8=75%).

TABLE 7.14 Effects of the treatment to fill the gaps for *Y/N-inversion* (Stage 4)

	Developmental stage on the pretest	Number of students who showed gaps on the pretest	Number of students whose gaps were filled
Treatment group	Stage 6 (N=22)	1	1
	Sage 5 (N=9)	0	-
	Stage 4 (N=4)	3	0
	Total	4	1 (25%)
Comparison group	Stage 6 (N=9)	0	-
	Stage 5 (N=3)	0	-
	Total	0	-

7.3.3.2 *Wh-inversion* (stage 4)

There were a total of 33 students (TG=26, CG=7) showing a gap for *Wh-inversion*. As discussed in Section 7.3.1, among the five rules for which a gap was found on the pretest, the largest proportion of students (70%) showed a gap for this rule (cf. *Y/N-inversion*: 4 of 47=9%, *Negative-Aux/Do-2nd*: 21 of 43=49%, *Cancel-inversion*: 18 of 31=58%, *Q-tag*: 13 of 31=42%). In the following analysis, I have excluded eight students' data (TG: Toyo, Nobu, Kaji, Yuta, Riko, Go, Jin; CG: Kiko) due to lack of obligatory occasions for applying this rule on at least one posttest. This brought the number of students showing the gap for this rule to 19 in the TG and six in the CG.

For the effect of the treatment on the acquisition of this rule, the picture is similar to that found for *Y/N-inversion*. In the TG, 19 students showed a gap for this rule (see Table 7.15 on the next page).

TABLE 7.15 Implicational scales for the students showing a gap for *Wh-inversion*

[Treatment group]

Stage 6		1	2	3	4wh	4yn	5wh	5Nwh	6C	6T	Gap filled
Ken	Pre	+	+	+	-	+	+	-	+	+	
Ken	Post	/	+	+	+	+	+	+	-	+	
Ken	Follow-up	+	+	+	-	+	+	+	+	+	
Yan	Pre	+	/	+	-	+	+	+	+	+	
Yan	Post	+	+	+	+	+	+	+	+	+	
Yan	Follow-up	/	+	+	-	+	-	-	+	+	
Ru	Pre	+	+	+	-	+	+	+	-	+	
Ru	Post	+	+	+	+	+	+	+	+	+	
Ru	Follow-up	+	+	+	-	+	+	+	+	-	
Hiro	Pre	+	+	+	-	+	+	-	-	+	
Hiro	Post	+	/	+	+	+	+	-	+	+	YES
Hiro	Follow-up	+	+	+	+	+	+	+	+	+	
Junko	Pre	+	+	+	-	+	+	+	-	+	
Junko	Post	+	+	+	+	+	+	-	+	+	YES
Junko	Follow-up	+	+	+	+	+	+	-	+	+	
Tomi	Pre	+	+	+	-	+	+	+	-	-	
Tomi	Post	+	+	+	-	+	+	+	+	+	
Tomi	Follow-up	+	/	+	+	+	+	+	+	-	
Fuku	Pre	+	+	+	-	+	+	-	+	-	
Fuku	Post	+	+	+	-	+	+	-	+	+	
Fuku	Follow-up	+	+	+	+	+	+	-	+	+	
Waka	Pre	+	+	+	-	+	+	-	-	-	
Waka	Post	+	+	+	-	+	+	+	+	+	
Waka	Follow-up	+	+	+	-	+	+	+	-	+	
Toby	Pre	+	+	+	-	+	+	+	-	+	
Toby	Post	+	+	+	-	+	+	+	+	+	
Toby	Follow-up	+	+	+	+	+	+	+	+	+	
Hide	Pre	+	+	+	-	+	+	+	-	+	
Hide	Post	/	+	+	-	+	+	+	-	+	
Hide	Follow-up	+	+	+	+	+	+	+	-	+	
Moto	Pretest	+	+	+	-	+	+	-	+	-	
Moto	Posttest	+	+	+	+	+	+	-	+	+	YES
Moto	ollow-up te	+	+	+	+	+	+	+	-	+	
Stage 5											
Jo	Pre	+	+	+	-	+	+	-	-	-	
Jo	Post	+	+	+	-	+	+	-	+	+	
Jo	Follow-up	+	+	+	-	+	+	-	+	+	
Nae	Pre	+	+	+	-	+	+	+	-	-	
Nae	Post	+	+	+	-	+	+	+	+	+	
Nae	Follow-up	+	+	+	+	+	+	-	+	+	
Yuya	Pre	+	+	+	-	+	+	-	-	-	
Yuya	Post	/	+	+	+	+	+	+	-	+	
Yuya	Follow-up	+	/	+	-	+	+	+	+	+	
Taka	Pre	+	+	+	-	+	+	+	-	-	
Taka	Post	+	+	+	-	+	+	+	-	+	
Taka	Follow-up	/	+	+	-	+	+	-	+	-	
Stage 4											
Omi	Pre	+	+	+	-	-	-	-	-	-	
Omi	Post	+	+	+	-	+	-	-	-	-	
Omi	Follow-up	+	+	+	-	-	-	-	-	+	
Rika	Pre	+	+	+	-	-	-	-	-	-	
Rika	Post	+	+	+	+	+	-	-	-	-	
Rika	Follow-up	+	+	+	-	-	-	-	-	-	
Kako	Pre	+	/	+	-	-	-	-	-	-	
Kako	Post	/	+	+	-	-	-	-	+	-	
Kako	Follow-up	+	+	+	-	-	-	-	-	-	
Mo	Pre	+	+	+	-	+	-	-	-	-	
Mo	Post	+	+	+	-	+	+	-	+	+	
Mo	Follow-up	+	+	+	+	+	+	+	+	-	

[Comparison group]

Stage 6		1	2	3	4wh	4yn	5wh	5Nwh	6C	6T	Gap filled
Taro	Pre	+	+	+	-	+	+	-	-	+	
Taro	Post	+	+	+	+	+	+	+	+	+	YES
Taro	Follow-up	+	+	+	+	+	+	-	+	+	
Dan	Pre	+	+	+	-	+	+	-	+	-	
Dan	Post	+	+	+	+	+	+	-	-	-	
Dan	Follow-up	+	+	+	-	+	+	-	+	+	
Yo	Pre	+	+	+	-	+	+	-	+	-	
Yo	Post	+	+	+	-	+	+	-	-	+	
Yo	Follow-up	+	+	+	-	+	+	-	-	-	
Stage 5											
Seto	Pre	+	+	+	-	+	+	-	-	-	
Seto	Post	+	+	+	+	+	+	-	-	-	YES
Seto	Follow-up	+	+	+	+	+	+	-	-	+	
Mi	Pre	+	+	+	-	+	+	+	-	-	
Mi	Post	+	+	+	-	+	+	-	-	-	
Mi	Follow-up	+	+	+	-	+	+	+	-	-	
Sae	Pre	+	+	+	-	+	+	-	-	-	
Sae	Post	+	+	+	-	+	+	-	-	+	
Sae	Follow-up	+	+	+	-	+	+	-	-	+	
Stage 4											

Their developmental stage on the pretest varied from stage 4 to 6 (four stage 4 students, four stage 5 students, and 11 stage 6 students). After the treatment, three of the 19 students (Hiro, Junko and Moto at stage 6) used two or more productive examples of this rule on both posttests (marked by a '+' for each posttest) and they were considered to have acquired *Wh-inversion*. The remaining 16 students were considered not to have acquired this rule, since five (Waka at stage 6, Jo and Taka at stage 5, and Omi and Kako at stage 4) got a '-' for both posttests and 11 (Ken, Yan, Ru, Tomi, Fuku, Toby and Hide at stage 6, Nae and Yuya at stage 5, and Rika and Mo at stage 4) gained a '+' only for one posttest. Of these 11 students being assigned a '+' for only one posttest, delayed acquisition was observed in six students (Mo, Nae, Tomi, Fuku, Toby, Hide).

In the CG, two of the six students (Seto at stage 5 and Taro at stage 6) produced two productive *Wh-inversion* questions on both posttests and they were considered to have acquired this rule. For the other four students, three students (Mi and Sae at stage 5 and Yo at stage 6) got a '-' for both posttests and one student (Dan at stage 6) gained a '+' only for the follow-up test. I summarise these results in Table 7.16 below.

TABLE 7.16 Effects of the treatment to fill the gaps for *Wh-inversion* (Stage 4)

	Developmental stage on the pretest	Number of students who showed gaps on the pretest	Number of students whose gaps were filled
Treatment group	Stage 6 (N=22)	11	3
	Stage 5 (N=9)	4	0
	Stage 4 (N=4)	4	0
	Total	19	3 (16%)
Comparison group	Stage 6 (N=9)	3	1
	Stage 5 (N=3)	3	1
	Total	6	2 (33%)

In the TG, the proportion of students considered to have acquired *Wh-inversion* was 16 percent (3 of 19), the smallest among the five rules for which a gap was observed (*Y/N-inversion*: 1 of 4=25%,

Negative-Aux/Do-2nd: 5 of 11=45%, *Cancel-inversion*: 9 of 11=82%, *Q-tag*: 6 of 8=75%). In the CG, the proportion of students considered to have acquired this rule was 33 percent (2 of 6), which was slightly larger among the four rules (*Negative-Aux/Do-2nd*: 1 of 8=13%, *Cancel-inversion*: 1 of 7=14%, *Q-tag*: 1 of 4=25%), but there was little difference between the two conditions. These results show that the treatment was not effective in helping students acquire *Wh-inversion* and the gap for this rule persisted in their production, and that mere task repetition by the CG students in the three tests was also not useful to help many of them acquire this rule. Their developmental stage on the pretest varied between stage 4 and stage 6, which indicates that they kept failing to apply *Wh-inversion*, even after reaching the highest level in L2 development in this area. Two reasons can be suggested for this.

The first is straightforward: many students did not acquire *Wh-inversion* through the treatment simply because it did not target this rule. Second, students' orientation towards communicative effectiveness during performing the information-gap tasks may be a reason for the frequent occurrence of the gaps for this rule. I noticed while analysing question examples in which students tried to apply *Wh-inversion* that many individuals, at different developmental levels, failed to place the copula in *wh*-questions and produced examples like 'How much the rent?' across the three tests. For instance, on the pretest a missing copula was observed with 17 of the 25 students showing a gap for *Wh-inversion* (three stage 4 students, six stage 5 students, and eight stage 6 students). As reviewed in Chapter 3, the copula is thought to be a variational feature, and its suppliance and non-suppliance is influenced by learners' orientation, such as their degree of norm-orientedness in using English, independent of their developmental stage. In other words, it is possible to say that many students' non-insertion of the copula in *wh*-questions in the three tests was caused by giving priority to communicative effectiveness over accuracy. In the information gaps tasks used in the tests, students were able to elicit information from the examiner, even if a copula was missing in their *wh*-questions, since it did not severely hinder the conveying of their message.

These observations suggest a need to focus on stage 4 Q-forms, even if learners have reached higher stages in L2 development. In many classroom tasks, learners use *Wh-inversion* to achieve immediate task goals and, if teachers are not aware of the problem caused by a missing copula and keep giving tasks without raising learners' attention to copula structure required in *Wh-inversion* questions, it may

lead learners to the use of simplified *wh*-questions without a copula and, worst of all, to fossilization.

7.4 Summary

I will now summarise the findings for the two research questions.

Q3. Do the students show developmental gaps in interrogative acquisition?

I have shown that many students across various developmental stages in both conditions revealed gaps in the acquisition of question formation, including gaps for Q-forms at their current developmental stage and/or for those at lower stage(s). Gaps were found for the following Q-forms: *Wh-inversion* (33 of 47 students=70%) and *Y/N-inversion* (4 of 47=9%), both at stage 4, *Negative-Aux/Do-2nd* at stage 5 (21 of 43=49%), and *Cancel-inversion* (18 of 31=58%) and *Q-tag* (13 of 31=42%), both at stage 6. No gap was observed for stage 5 *Affirmative-Aux/Do-2nd*.

Q4. Is the instruction effective in filling such gaps?

I have demonstrated that the treatment was effective in filling the gaps for the target Q-forms. Table 7.17 below summarises the results.

TABLE 7.17 · Percentage of students whose gap was filled

		Treatment group		Comparison group	
Stage 6	*Cancel-inversion*	9/11	(82%)	1/7	(14%)
	Q-tag	7/ 9	(78%)	1/4	(25%)
Stage 5	*Negative-Aux/Do-2nd*	5/11	(45%)	1/8	(13%)
	Affirmative-Aux/Do-2nd	-	-	-	-
Stage 4	*Y/N-inversion*	1/ 4	(25%)	-	-
	Wh-inversion	3/19	(16%)	2/6	(33%)

For the target rules, in the TG the proportion of students whose gap was filled through the treatment was: 82 percent (9 of 11) for stage 6 *Cancel-inversion*, 78 percent (7 of 9) for stage 6 *Q-tag*, and 45 percent (5 of 11) for stage 5 *Negative-Aux/Do-2nd*. On the other hand, in the CG the percentage of students whose gap was filled was much smaller: 14 percent (1 of 7) for stage 6 *Cancel-inversion*, 25 percent (1 of 4) for stage 6 *Q-tag*, and 13 percent (1 of 8) for stage 5 *Negative-Aux/Do-2nd*. The larger proportion observed in the TG clearly show the effectiveness of the treatment to help students acquire the target rules.

As for untargeted two stage 4 rules, in both conditions the proportion of students whose gap was filled on the posttests was small and little difference was observed between the two conditions. For example, the proportion of students whose gap for *Wh-inversion* was filled was 16 percent (3 of 19) in the TG, while in the CG it was 33 percent (2 of 6), and for *Y/N-inversion* in the TG it was 25 percent (1 of 4). I explained the limited effect of the treatment by giving two reasons. One was that the treatment did not target these stage 4 Q-forms. The other reason for the persistent gap was learners' orientation toward communicative effectiveness and the lack of sensitivity to copula structure. I pointed out that stage 4 Q-forms involve copula 'be', one of the variational features, and directed teachers' attention to these problems. I have suggested the need for drawing learners' attention to their simplified usage of stage 4 questions, especially *Wh-inversion* questions.

CHAPTER 8
Learners' perceptions of the usefulness of the treatment

In the previous two chapters, I have analysed student output and shown that the noticing-promotion approach was effective in helping the Ready students not only to advance in developmental stage in their acquisition of question formation, but to fill the gaps in their acquisition of the target rules. These findings are encouraging for teachers who teach large classes, where immediate teacher feedback is difficult to provide, but the product data analysis has not shown how the students perceived the usefulness of the treatment. I should remind the reader that Lynch and Maclean (2000, 2001) reported that learners' L2 level was closely linked with their awareness both of changes made in the repetition of a meaning-focused conversation task called Poster Carousel, which was given without a specific focus on form, and of improvements in their successive performances. While higher-level students were aware of their gains and consciously planned changes to their English during the task, those at lower levels were unaware of their improvement in various areas of English usage. This may be the case in this study, although the tasks in this study were designed to draw the students' attention to *form*.

Lynch and Maclean (2001) also point out that many L2 learners are unaware of the availability and longer-term value of negotiated input and output with other non-native speakers in the classroom. The noticing-promotion approach required the students to take an active role in assisting peer learner while working in pairs most of the time, and it was hoped that this requirement would help learners to notice the availability and value of peer interaction and to appreciate what they stand to gain from conversations with peer learners. However, given my involvement both as task designer and teacher, I was concerned about my students' own views on the value of practice. In order to explore the students' perceptions of the usefulness of the instruction and of their changes in producing target questions inside and outside of L2 classes, and in their classroom behaviours in pair work, I formulated the following five research questions:

Q5. Do learners think the modelling video they have watched is useful in understanding how to carry out the subsequent dictogloss?

Q6. Do the students think the modelling video is useful regarding any other aspects of L2 learning?

Q7. Are there any differences in the noticing (i.e. reported successful self-corrections) of Ready students in the subsequent tasks depending on the type of modelling video they have watched?

Q8. Are there any differences in the noticing (i.e. reported successful self-corrections) of Unready students in the subsequent tasks depending on the type of modelling video they have watched?

Q9. Do the students notice any positive changes in their use of the target rules beyond the treatment sessions?

Q10. If so, what kinds of changes have they noticed?

In Section 8.1 I will examine their rating of the usefulness of the modelling video in the questionnaire, which was used as a means of showing the students how to work together in the subsequent question reconstruction task; then I shall examine their assessment of its benefit for other aspects of L2 learning. In Section 8.2 I will examine the reports on partner-assisted successful self-corrections in the conversation tasks given after watching the modelling video. It would be ideal to analyse the students' classroom interaction along with the questionnaires, but that would involve further transcribing and is beyond the scope of this study. As an alternative, I will quote some concrete examples of partner-assisted successful self-corrections reported in the questionnaire given after the first treatment session. In Section 8.3 I will examine their self-reports, collected in the questionnaire given after the follow-up test, regarding positive changes in their use of the target Q-forms in contexts beyond the treatment sessions. My interest is in seeing whether the instruction was beneficial both in encouraging the students to scaffold each other for the acquisition of the target rules in pair work, and in helping them carry on using those Q-forms for real communication outside class, since this is the ultimate goal of L2 teaching.

8.1 Usefulness of the modelling video as input

I will now examine from two perspectives the students' perceptions of the usefulness of watching the modelling video: (a) their rating of its usefulness in helping them understand how to work together to reconstruct questions in the subsequent dictogloss task and (b) their reports on its benefits for other aspects of L2 learning. Before presenting my analysis of the questionnaire, I will briefly review the content of the noticing-promotion approach, in order to remind the reader how the modelling video was used.

As explained in Chapter 5, I taught the students the target rule explicitly in the mini-lecture at the beginning of each treatment session (for the flow of the instruction, see Figure 5.1 on page 104). For instance, in Week 1 the target rule was *Q-tag* and I taught the usage of agreement and intonation in the tag question by using a handout. After the mini-lesson, I gave the student dyads two sets of dictogloss task. Each dictogloss consisted of two parts: (a) listening to the conversation between two native English speakers recorded on the audio-tape and (b) reconstructing the question examples in pairs by using notes from the preceding listening. During listening, although they were allowed to look at the notes while reconstructing the questions, they had to use their knowledge of the target grammar, since the conversation was recorded at natural speed and they did not have time to write down all of each question.

While one student reconstructed half the questions using their own notes, the other student wrote down the reconstructed questions on an answer sheet provided, while checking whether a Q-form was appropriately applied in each case. Halfway through the reconstruction task, they changed roles. Before starting the reconstruction task, the students were encouraged to pay attention to their partner's output and to take an active role in assisting their partner to notice their mistakes and make self-corrections. In the first dictogloss, after listening to the recorded conversation, the students watched the modelling video which showed how to do the reconstruction task. The video was also intended to provide them with useful words and expressions necessary to talk about the grammar in English, as discussed in Chapter 5. The student dyads were then given the information-gap task intended to provide opportunities to apply the target rule in more demanding conditions than in the

dictogloss.

I should also remind the reader that two types of modelling video were used in each treatment session (the self-correction version and peer-assistance version) in order to examine whether the inclusion of the scenes of peer assistance could promote active interaction and noticing in the subsequent task. As explained in Chapter 5, one of the two classes at each institution watched the same version throughout the four treatment sessions, and the other class watched the other version. In both versions, the same student dyad worked cooperatively to reconstruct the same number of questions, but there were differences in some scenes in the way self-corrections were made. For example, in the two versions of the modelling video used in Week 1, in reconstructing 10 tag questions in the Memory Game, the student dyad made eight mistakes in total. In the self-correction version, all mistakes were self-corrected *without* peer assistance, while in the peer-assistance version, six of the eight mistakes were self-corrected *with* partner assistance, while the other two were self-corrected *without* peer assistance, as in the other version.

My prediction was that the students who had watched the concrete examples of collaboration in the peer-assistance version would try to provide their partner with assistance more frequently than those who had watched the self-correction version. I hoped that the resultant frequent experience of giving and receiving peer assistance in the subsequent tasks would increase their chances of not only noticing but also acquiring the target rule. I should, however, point out that no substantial differences were found in their acquisition of Q-forms between those who had watched the peer-assistance version and those who watched the other version, as discussed in Chapters 6 and 7. This result was brought about partly by the very small sample size. For instance, of the 18 students who watched the peer-assistance version, only three were in fact ready to learn stage 6 rules and the other 15 were unready, so meaningful comparison between the two conditions in the light of their readiness was not feasible. Thus, it was not possible to examine the effects of the type of modelling video on the students' acquisition of Q-forms, but analysis of their reports in the questionnaire has revealed that the type of modelling video did in fact have an influence on their classroom experiences.

I will now present, in turn, the results of my analysis to address the aforementioned two research

questions (i.e. Q5 and Q 6).

8.1.1 Do learners think the modelling video they have watched is useful in understanding how to carry out the subsequent pair work in the dictogloss? (Q5)

8.1.1.1 Procedures for the analysis

In the questionnaire completed immediately after the class, using a five-point scale, where *1* was 'not useful at all' and *5* was 'very useful', each student rated the usefulness of the modelling video in understanding how to carry out the dictogloss reconstruction task in pairs. I counted the number of students who gave each rating in relation to their readiness to learn stage 6 rules and the type of video they had watched. In the analysis, I categorised the students whose developmental stage was 4 or below on the pretest as Unready, and assigned them either to the Unready-Peer group or the Unready-Self group according to the type of video they had watched. For the other students whose developmental stage on the pretest was stage 5 or stage 6, I categorised them as 'Ready' and assigned them to either the Ready-Peer group or the Ready-Self group. I should point out that in the previous two chapters the students at stage 6 were not categorised as 'Ready' but as 'Stage 6 students', but in the following analysis I treat them as 'Ready'. This is because, as we have seen in Chapter 7, many of them (15 of 22) had not acquired one or two stage 6 rule(s) and they were still, in a sense, in the process of learning them; indeed they were ready to learn. The results of this group assignment are shown in Table 8.1 below.

TABLE 8.1 Number of students who watched each version of modelling video

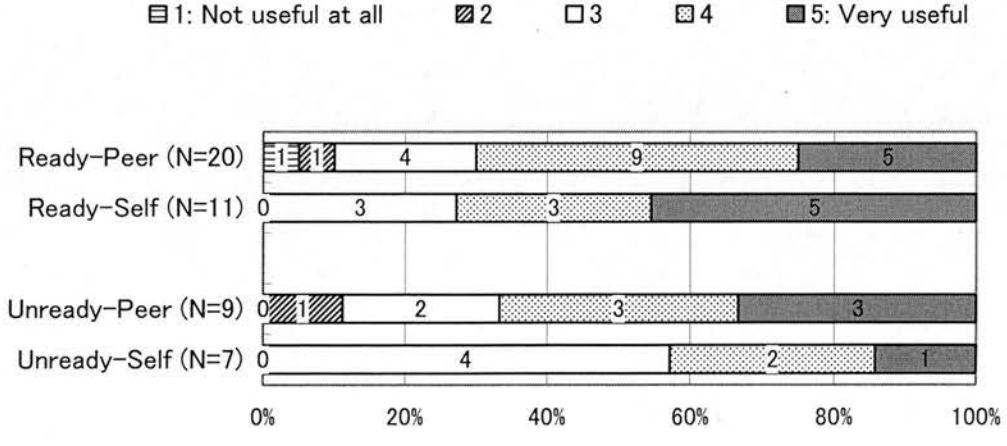
Group	Developmental stage on pretest					
	Stage 1	Stage 2	Stage 3	Stage 4	Stage 5	Stage 6
Ready-Peer (N=20)					6	14
Ready-Self (N=11)					3	8
Unready-Peer (N=9)	0	0	6	3		
Unready-Self (N=7)	1	0	5	1		

For the two Ready groups, the students who watched the peer-assistance version (Ready-Peer) comprised 20 students, including six stage 5 students (30%) and 14 stage 6 students (70%). The other Ready group who watched the self-correction version (Ready-Self) consisted of 11 students, including three stage 5 students (27%) and eight stage 6 students (73%). In each Ready group, the proportion at each developmental stage was almost the same. On the other hand, for the two Unready groups, those who watched the peer-assistance version (Unready-Peer) comprised nine students, including six stage 3 students (44%) and three stage 4 students (56%). The other Unready group who watched the self-correction version (Unready-Self) consisted of seven students, including one stage 1 student (14%), five stage 3 students (57%) and one stage 4 student (29%); and the average of their developmental stage was lower than that of the other Unready group. This means that the general English proficiency of the students in the Unready-Self group was lower than those in the Unready-Peer group.

8.1.1.2 Findings and discussion

I now turn to the results of the students' rating of the usefulness of each version of modelling video for their understanding of the question reconstruction task given after watching the modelling. As space is limited, I will confine my examination to the data collected in the questionnaire given in Week 1. The number of students who gave each rating is shown in Figure 8.1 below.

FIGURE 8.1 Usefulness of the modelling video in understanding how to do the reconstruction task (Week 1)



For the two Ready groups, a similar proportion of students gave a rating of *4* or *5*. For instance, this high rating was given by 14 of the 20 students (70%) in the Ready-Peer and by eight of the 11 students (73%) in the Ready-Self. For the other students in the Ready-Peer, *3* was given by four students and *2* or *1* by one student for each. For the Ready-Self, the remaining three students gave *3*. These results suggest that many Ready students benefited from the modelling video, irrespective of the type of modelling they had watched.

For the two Unready groups, the proportion rating *4* or *5* was six of the nine students (67%) in the Unready-Peer and three of the seven students (43%) in the Unready-Self. Of the other students in the Unready-Peer one student gave *2* and two students gave *3*, while in the Unready-Self the rest of all four students gave *3*. The proportion giving a high rating in the two Unready groups was larger in the Unready-Peer and two reasons can be suggested for the smaller proportion giving a high rating in the Unready-Self. The first is straightforward: the average proficiency of the students in the Unready-Self group was lower than that in the Unready-Peer group, and it is not difficult to imagine that more students in the former group experienced problems in understanding the conversation held exclusively in English at near-natural speed. The second reason involves a difference in the type of modelling each group watched. In the peer-assistance version, the interaction in the student dyad was more active than in the other version, since it included the scenes of peer assistance. It is conceivable this active interaction made the content of the conversation more salient and increased their understanding of the task procedures.

Thus, the students' readiness to learn the target rule affected their understanding of the modelling which contained the examples of target rules, but it is interesting to point out that the proportion giving a high rating in the Unready-Peer (67%) was as large as in the two Ready groups (Peer=70%, Self=73%). This suggests the peer-assistance modelling video was more effective in helping many Ready and Unready students at different developmental stages understand how to do the subsequent dictogloss reconstruction task in pairs.

I will now turn to examine the students' reports on the benefit of the modelling video for other aspects of L2 learning.

8.1.2 Do the students think they have benefited from the modelling video regarding any other aspects of L2 learning? (Q6)

In the same questionnaire, the students were asked whether they had benefited from the modelling video in other aspects of L2 learning (i.e. apart from learning how to do the subsequent task). It was an open question, and so some students answered the question and others did not. The main purpose of the video was to help the students understand the task procedures, but it was also intended to provide further input, such as 'the usage of target rule' and 'the non-target L2 items' in both versions, and 'effective ways of assisting the partner to notice their own mistakes and make self-corrections' in the peer-assistance version. It is interesting to see to what extent the students' reports differ according to their readiness to learn the target rule and the type of modelling they watched. I will first examine the number of students who reported benefits in the four conditions, and then, the content of their reports.

8.1.2.1 Number of students who reported a benefit

The number of students who reported a benefit in each condition is summarised in Table 8.2 below.

TABLE 8.2 Number of comments on the usefulness of the modelling video other than understanding how to carry out the task

		Number of students	
Ready-Peer group	(N=20)	13	(65%)
Ready-Self group	(N=11)	9	(82%)
Unready-Peer group	(N=9)	7	(78%)
Unready-Self group	(N=7)	4	(57%)

For the two Ready groups, reports were made by 13 of the 20 students (65%) in the Ready-Peer and by nine of the 11 students (82%) in the Ready-Self. For the two Unready groups, seven of the nine students (78%) in the Unready-Peer and four of the seven students (57%) in the Unready-Self commented on its benefit. Again, the proportion of students reporting a benefit was smaller in the Unready-Self than in the other three conditions, but it is noteworthy that more than half of the students at

different developmental stages in all conditions answered that they had benefited from watching each version of the modelling video. The main purpose of the modelling was to show the students how to carry out the task in pairs, and the reported experience of benefit by these students was derived from the additional effects of the seeded input in the video. One can say that both versions served as a useful means of providing the students at different developmental stages with L2 input.

I will now examine in detail the content of the reported benefits, in relation to the type of modelling video the students had watched, and their readiness to learn the target rule.

8.1.2.2 Content of the reported benefits

I will first examine the reports made by the two groups who watched the peer-assistance version, and then those by the two groups who watched the self-correction version.

8.1.2.2.1 Benefits of the peer-assistance version

One Ready group and one Unready group watched the peer-assistance version. As noted earlier, in this version the following three kinds of items were seeded as *input*:

(a) 'effective ways of assisting the partner'
(b) 'the target rule (*Q-tag*)'
(c) 'the non-target L2 items'

The reported benefits in each group fell into these categories and I summarise them in Table 8.3 on the next page. I will, in turn, examine the reported benefits in each category of the seeded input.

TABLE 8.3 Number of students reporting benefits of the two types of modelling video (Week 1)

	Type of modelling video			
	Peer-assistance		*Self-correction*	
	Ready (N=20)	Unready (N=9)	Ready (N=11)	Unready (N=7)
Effective ways of assisting the partner	5 (25%)	6 (67%)	0 (0%)	0 (0%)
Target Q-form (*Q-tag*)	6 (30%)	1 (11%)	0 (0%)	1 (14%)
Agreement		1		1
Intonation to express certainty of the statement	3	1		
Technical terms to talk about the target grammar	2			
Noticing possible mistakes	2			
Non-target L2 items	5 (25%)	5 (56%)	9 (82%)	4 (57%)
Useful expressions for the management of conversation	3	2	4	1
How to use fillers		2	2	
Pronunciation		4	2	2
Other	2		2	2

Effective ways of assisting the partner

Benefits involving 'effective ways of assisting the partner' were reported by five students (25%) in the Ready-Peer and by six of the nine students (67%) in the Unready-Peer. The following two extracts from the questionnaire illustrate the content of the reported benefits.

> '(I learnt) how I could help my partner. For example, giving the partner time to reflect on her utterance to make self-corrections is a good way, instead of telling the partner that 'you've made a mistake'.' ---Waka at stage 6

'(I learnt) how to tell the partner that the cause of mistakes was intonation in tag questions' --- Taka at stage 5

The peer-assistance version of modelling included such scenes as giving the partner time to reflect on his/her own output and providing hints as to mistakes in the output. The above-cited reports show that the seeded input in the video was taken up by those students. I should remind the reader that I use the term *input* to refer to more than linguistic input. The modelling video was a means of providing the students with a rich variety of 'affordances' (van Lier, 1996:53) - learning opportunities through conversational interaction in the classroom. The students' reported experiences are good evidence that the use of the peer-assistance version of the modelling video was effective in providing the students with affordances of the sort van Lier has argued for.

One possible reason for the lower proportion in the Ready-Peer could be their familiarity with the ways in which the assistance was offered in the modelling, such as repeating a partner's erroneous utterance to signal a mistake and explaining the target grammar. Arguably, giving such assistance in L2 learning was not novel for many Ready students at a higher developmental stage (i.e. stage 5 or stage 6), and for many of them it was not worth reporting as a benefit. On the other hand, for many Unready students watching the concrete examples of offering such assistance seemed a useful experience, since L2 learners are seldom given opportunities to watch this sort of modelled peer assistance. The following two comments given by two Unready students at stage 4 or stage 3 were good examples of such learning.

'(I learnt) how to offer assistance when the partner made mistakes or stumbled at words.'
by Tomo at stage 3

'(I learnt) how to give my partner hints when she made mistakes'
by Chie at stage 3

The target rule (*Q-tag*)

While six of the 20 students (30%) in the Ready-Peer reported that the modelling video was useful to learn 'the target rule (*Q-tag*)', with one exception, the majority of students in the Unready-Peer (8 of

9=89%) did not. The lack of reports in the Unready-Peer suggests that the peer-assistance version was not effective in providing many Unready students with the input to learn *Q-tag*. It is not difficult to imagine that many Unready students had problems in understanding the dialogue including the examples of target rule and the metatalk about its usage, since they were not ready to learn the target rule.

The benefits reported by the six Ready-Peer students concerned the following three items: (a) 'use of intonation to express a different degree of certainty of the statement in tag questions' (3 students), (b) 'technical terms to talk about the target grammar' (2 students), and (c) 'noticing possible mistakes in producing tag questions in the subsequent tasks' (2 students). The report by the one Unready-Peer student involved 'agreement' and 'intonation' in tag questions.

The non-target L2 items

Benefits involving 'the non-target L2 items' were reported by five of the 20 students (25%) in the Ready-Peer and the reports involved 'useful expressions for the management of conversation' (3 students) and 'learning importance of responding to the partner in a lively way to enhance learning' (2 students). In the Unready-Peer, five of the nine students (56%) made comments on 'the non-target L2 items' and their comments involved 'useful expressions for the management of conversation' (2 students), 'how to use fillers' (2 students) and 'pronunciation' (4 students). The smaller proportion of students reporting the benefits in the Ready-Peer can be explained by the Ready students' prior knowledge of the seeded input. It is probable that they were able to use fillers in English and knew the pronunciation of words used in the modelling. In fact, none of the 20 Ready mentioned the use of fillers and pronunciation as a benefit.

I will now turn to examine the benefits reported by the other two groups, who watched the self-correction version.

8.1.2.2.2 Benefits of the self-correction version

One Ready group (N=11) and one Unready group (N=7) watched the self-correction version, and a similar trend was observed in the content of their reports on its benefits. For instance, no student in either group made a report concerning 'effective ways of assisting the partner' because the modelling they had watched did not include scenes of peer assistance, although the student dyad on the video cooperated with each other to complete the reconstruction task. This suggests a need to show scenes of peer assistance in modelling to provide learners with opportunities to learn how they can be effective in assisting their partner. As we have seen, in the two groups who had watched the other version, both the Ready and Unready students reported its benefits, which suggests that the modelling video performed by the student dyad can be a useful tool if teachers wish learners to take an active role in peer interaction

In the L2 class, modelling is usually given by the teacher in person, and learners tend to have few chances to see peer learners making self-corrections with the assistance from the partner. This has the merit that teachers can demonstrate the modelling while checking learners' understanding and can slow the speed of talk where necessary - two advantages not available in the modelling video; however, modelling given by the teacher has disadvantages, too. For instance, it can be intimidating to learners if the teacher uses examples of their mistakes in on-line conversation to teach how to give feedback on the partner's mistake. Such problems can be avoided if a recorded conversation by an unknown student pair is used, as in the modelling video used in this study. In the treatment, before showing the video, I told the students that 'you will see the student dyad making mistakes, but you can learn lots of things from their performance since they were good at noticing their mistakes and cooperating with each other'. From the outset, the students looked interested in watching peer learners' performance on the video, but having been told this, they seemed to become more interested in its content[36].

I will now examine the reported benefit in the two groups.

The target rule (Q-tag)

None of the 11 students in the Ready-Self reported a benefit involving 'the usage of target rule', despite their readiness to learn it. The fact that 30 percent of students in the Ready-Peer (6 of 20 students) reported this benefit suggests that the self-correction version was not as effective as the other in this respect. In the self-correction version, all mistakes - including ones in applying the target rule - were self-corrected without needing any assistance, and it is conceivable that the most salient thing in the modelling was the smooth interaction between two competent students able to use various expressions to keep the conversation going while correcting the mistakes by themselves.

With one exception, the students in the Unready-Self (6 of 7) did not report on this item, either. As we have seen earlier, in the other Unready group who had had the benefit of watching the peer-assistance version, only one student made a report involving 'the target rule' and the lack of reports in both Unready groups can be explained by their unreadiness to learn stage 6 *Q-tag*.

The non-target L2 items

In the Ready-Self, benefit involving 'the non-target L2 items' was reported by nine of the 11 students (82%), including the reports on 'the use of useful expressions for the management of conversation' (4 students), 'use of fillers' (2 students) and 'pronunciation' (2 students). Given that the proportion of students in the other Ready group who had watched the other version was smaller (5 of 20=25%), it is possible to conclude that the self-correction version was more effective in providing input involving 'the non-target L2 items'.

In the Unready-Self, four of the seven students (57%) reported benefits on this, referring to 'pronunciation' (2 students), 'useful expressions to carry out the task' (1 student), and 'others' including 'facial expressions' (2 students). We have seen in the previous section (Section 8.1.2.2.2) that a similar proportion of students (55%) in the other Unready group who had watched the peer-assistance version made reports on 'the non-target L2 items'. This shows that the type of modelling video produced no differential effects in the two Unready groups, irrespective of the type of modelling they had watched.

As for the finding of the difference in the proportion of students reporting the benefits of 'the non-target L2 items' in the two groups who had watched the self-correction version (Ready-Self=82%, Unready-Self=57%), this modelling video was not seeded with 'effective ways of providing peer assistance' but with 'the usage of target rule' and 'the non-target L2 items'. Two possible reasons for this finding can be suggested. One is the students' readiness: the lower level of English proficiency in the Unready-Self seemed to limit their intake from the seeded input. It is also conceivable that they were engaged in ensuring they understood the task procedure shown on the video and were not able to pay attention to the seeded input involving 'the non-target L2 items'.

8.1.2.3 Summary for Section 8.1

We have seen both the potential and the limitations of using the two versions of modelling video as a means of providing the input in L2 learning. I will summarise the findings for the two research questions.

Q5. Do learners think the modelling video they have watched is useful in understanding how to carry out the subsequent pair work in the dictogloss?

One can say that both versions of the video, to a large extent, served the original purpose as input to help many students at different developmental stages understand how to carry out the subsequent reconstruction task in pairs. We have seen that about 70 percent of students gave the high rating of *4* or *5* in three of the four conditions, irrespective of the types of modelling they had watched (Ready-Peer=14 of 20 students, Unready-Peer=6 of 9, Ready-Self=8 of 11). Most of the students in each condition awarded *3*. In the Unready-Self, whose average developmental stage was lower than the other Unready group, the proportion giving such high rating was lowest (3 of 7=43%) and I have attributed this to a combination of their unreadiness to learn the target rule and the type of modelling video they had watched.

Q6. Do the students think they have benefited from the modelling video regarding any other aspects of L2 learning?

We have seen that both Ready and Unready students commented on the benefit of each type of modelling video and that the proportion reporting the benefit differed, depending on the type of modelling video they had watched and/or their readiness to learn the target Q-form. For example, the content of self-reports made in the two groups watching the peer-assistance version of the modelling video varied in relation to three aspects of L2 learning: (a) 'effective ways of assisting the partner', (b) 'the target Q-form', and (c) 'the non-target L2 items', while that in the other two groups who had watched the self-correction version involved the latter two aspects only.

As for 'effective ways of assisting the partner', unsurprisingly this benefit was reported only by those who had watched the peer-assistance version (5 of 20 Ready=25%, 6 of 9 Unready=67%). I have pointed out the need to include scenes of partner assistance to teach learners how they can offer useful help in pair work.

For the use of the modelling video as a means of providing L2 input concerning 'the target Q-form', the peer-assistance version was more effective than the other, and the Ready students benefited more from the peer-assistance version than the Unready students. As we have seen, the reports on the benefit of this were most frequently made in the Ready-Peer (6 of 20 students=30%) and their reports mentioned 'intonation to express certainty of the statement in tag questions', technical terms to talk about the target rule', and 'noticing possible mistakes in producing tag questions'. As for the other three groups, not a single report was made in the Ready-Self, despite their readiness to learn the target rule (Q-tag), and only one student commented on this benefit in the two Unready groups. On the absence of report in one of the two Ready groups (i.e. Ready-Self), I have argued that this was brought about by the difference in the types of modelling video; in the self-correction version the most salient feature was the smooth conversation, while the peer-assistance version featured active interaction to help the partner notice and self-correct mistakes; it may well be that this enhanced the saliency of the seeded input. Bearing in mind the limitations of using self-reports taken from a small sample, I concluded that the peer-assistance version was more effective for the Ready students to this end. However, the proportion reporting the benefit involving 'the target Q-form' in the two Ready groups (Peer=30%, Self=0%) was lower than I

had expected and I have attributed this to many Ready students' familiarity with the target rule. In fact, more than two thirds of the students in each group had reached stage 6 and many of them had acquired Q-tag before the treatment. It is reasonable to infer that they thought the seeded input involving 'agreement' and 'intonation' in tag questions was not worth reporting as a benefit. I have also argued that the lower proportion reporting this benefit in the two Unready groups was due to their unreadiness to learn the target rule and that it prevented them from understanding the recorded conversation in reconstructing the 10 examples of target rule (*Q-tag*).

In terms of the use of the modelling video as a means of providing the L2 input concerning 'the non-target L2 items', the amount of input included in each version and the students' English proficiency had effects on the proportion of students reporting the benefit. For instance, we have seen a difference between the two Ready groups; the proportion reporting this benefit was larger in the Ready-Peer (9 of 11=82%) than in the Ready-Self (5 of 20=25%). I have argued that the difference was brought about by the types of modelling video they had watched; the peer-assistance version was richer in terms of L2 learning content because of the inclusion of peer assistance and it was intended to draw their attention more to the seeded input concerning 'effective ways of assisting the partner' and the target Q-form'. In the two Unready groups, more than half the students in each made a report on 'the benefit involving 'the non-target L2 items'(5 of 9 Unready-Peer students=56%; 4 of 7 Unready-Self students=57%). Given the much lower proportion reporting the benefit concerning 'the target rule' in these groups (i.e. only one student in each made a report), the higher proportion reporting the benefit involving 'the non-target L2 items' in both groups suggests that what learners can learn from the modelling is what they are ready to learn. These findings have clear implications for teaching. I have pointed out the importance of teachers' making selective decisions as to what to include in the modelling in order to maximise learners' intake from the seeded input.

8.2 Successful self-corrections made *with* peer assistance

Throughout the four treatment sessions, to promote noticing and a cooperative relationship, I encouraged the students to do two things: (a) to pay attention to their own output so as to notice mistakes and (b) to take an active role in assisting each other in pair work. I was aware that learners' limited attentional resources would not allow them to pay attention to the output all the time while performing the tasks, but I hoped this encouragement would facilitate their noticing and cooperation in pair work. As this, in turn, should enhance learning, I will now examine its effects.

In the questionnaire given in Week 1 (see Appendix 25), I asked the students to report their own mistakes in using the target rule (*Q-tag*) in the conversation tasks (Q7), and then, in the following two questions, to tick the mistakes they had corrected by themselves *without* partner assistance (Q8) along with those corrected *with* partner assistance (Q9). I will first analyse the self-reports made by the two Ready groups, in relation to their readiness to learn the target rule (*Q-tag*) and the type of modelling video they had watched, and then I will analyse those by the two Unready groups.

8.2.1 The two Ready groups

To this end, the following research question was formulated , as outlined earlier.

Q7. Are there any differences in the noticing (i.e. reported successful self-corrections) of Ready students in the subsequent tasks depending on the type of modelling video they have watched?

The two Ready groups watched either the peer-assistance or the self-correction version. I will first examine the number of students in each group who reported successful self-corrections made *with* or *without* partner assistance. Then, I will examine whether there were any differences between the two Ready groups in the content of the reported self-corrections.

8.2.1.1 Number of reported successful self-corrections (Ready)

I counted the number of students in each condition who reported their self-corrections involving 'the target rule' and 'non-target L2 items' separately, *with* or *without* peer assistance, and I summarised its results in Table 8.4 below. I will first examine the content of self-corrections made *without* peer assistance.

TABLE 8.4 Number of students reporting self-corrections (Ready)

	Number of students reported own self-corrections	
	Target rule (Q-tag)	Non-target L2 items
***Without* assistance**		
Ready-Peer group (N=20)	8 (40%)	3 (15%)
Ready-Self group (N=11)	4 (36%)	3 (27%)
***With* assistance**		
Ready-Peer group (N=20)	12 (60%)	5 (25%)
Ready-Self group (N=11)	3 (27%)	2 (18%)

Self-corrections *without* peer assistance

For the reports concerning the target rule, the number of students making a positive report in the two Ready groups was similar: eight of the 20 students (40%) who had watched the peer-assistance version and four of the 11 students (36%) who had watched the self-correction version. For the reports concerning 'non-target L2 items', three students in each Ready group reported self-corrections made *without* assistance (Peer=15%, Self=27%) and the proportion was slightly larger among those watching the self-correction version. These results show that in each Ready group, the students made the corrections by themselves *without* help from the partner, and that the proportion of students who made the corrections involving 'the target rule' was larger than that concerning 'non-target L2 items'. On the

whole, no major difference was observed between the two conditions, which suggests that little difference was accounted for by the types of modelling video they had watched.

Self-corrections *with* peer assistance

I will now examine the number of students who reported partner-assisted self-corrections. In the two Ready groups, a similar trend was observed in the proportion of students reporting successful self-corrections concerning 'the non-target L2 items'. The number of students reporting successful self-corrections was five of the 20 (25%) in the Ready-Peer and two of the 11 (18%) in the Ready-Self.

On the other hand, for the reported successful self-corrections involving 'the target rule', a difference was observed between the two conditions. The number of students reporting self-corrections *with* partner assistance, as they had seen in the modelling video, was 12 of the 20 students (60%) in the Ready-Peer. In contrast, in the Ready-Self the number of students reporting this was only three of 11 (27%). Given that a similar proportion of students in each condition reported self-corrections concerning 'the target rule' made *without* assistance (Ready-Peer=40%, Ready-Self=36%), this difference must have been due to the types of modelling they had watched. Unsurprisingly, the peer-assistance version was more effective in promoting partner assistance, which led the peer learner to make self-corrections.

I should, however, point out that my discussion has been based on the students' self-reports on their classroom experience. There is a possibility that the reported assistance did not actually lead them to make successful self-corrections. It is also possible that, although they filled it in immediately after the class, some students might have forgotten the self-corrections they had made and failed to report them in the questionnaire. Indeed, this is a limitation of using learners' self-reports on classroom experiences without analysing their classroom interaction to support them. However, it is still worth examining the self-reports, since they embody the students' perceptions of the instruction's usefulness, which are closely linked with their satisfaction with their achievement in the class, as well as with their learning experience in pairs.

Bearing in mind the limitations of using the self-reports, I will now examine in further detail whether the type of modelling had an influence on the content of reported successful self-corrections.

8.2.1.2 The content of the self-corrections (Ready)

I will first examine the reported self-corrections made *without* partner assistance, and then, those *with* partner assistance.

Without assistance

For the self-corrections made *without* assistance by the Ready students, we have seen that the type of modelling did not make a difference to the proportion commenting on successful self-corrections in using 'the target rule' (Ready-Peer=40%, Ready-Self=36%). As for their content, a similar trend was observed between the two groups, as shown in Table 8.5 on the next page.

For example, in each group the reported successful self-corrections involved three kinds of mistake in using the target rule: 'agreement' (Peer=4 students, Self=2 students), 'intonation' (Peer=4 students, Self=3 students), and 'how to answer tag questions' (Peer and Self=1 student for each).

For the reported successful self-corrections involving 'the non-target L2 items', the picture was similar. As we have seen, the proportion reporting this was similar in the two groups (3 of 20 Ready-Peer students=15%, 3 of 11 Ready-Self students=27%), and the students in both groups mentioned similar mistakes, such as the misuse of articles.

With assistance

A similar trend was found for the reported self-corrections made *with* partner assistance involving 'non-target L2 items'. The proportion reporting this positive experience was similar: three of the 20 students (15%) in the Ready-Peer and two of the 11 students (18%) in the Peer-Self, and the reported self-corrections in both groups involved similar mistakes, such as articles.

TABLE 8.5 Number of self-reports on successful self-corrections in the two Ready groups (Week 1)

	***Without* assistance**		***With* assistance**	
	Ready-Peer (N=20)	Ready-Self (N=11)	Ready-Peer (N=20)	Ready-Self (N=11)
Target Q-form (*Q-tag*)				
● Agreement (negative-positive/ positive-negative)	4 (20%)	2 (18%)	9 (45%)	
● Intonation to express certainty of the statement	4 (20%)	3 (27%)	3 (15%)	2 (18%)
How to answer tag questions	1 (5%)	1 (9%)	2 (10%)	1 (9%)
Use of interrogative in the main clause			1 (5%)	
Non-target L2 items	3 (15%)	3 (27%)	3 (15%)	2 (18%)

Note. The items marked by ● were dealt with in both versions of modelling video as mistakes.

As to the reported successful self-corrections concerning 'the target rule' (*Q-tag*), we have seen that the proportion reporting the successful self-correction made *with* assistance was larger in the Ready-Peer (60%) than in the Ready-Self (27%). I examined the details of the self-corrections and found differences between the two Ready groups. For instance, as shown in Table 8.5, the reported self-corrections *with* assistance by the Ready-Peer ranged over four kinds of mistake, while those by the Ready-Self involved only two out of four. The four kinds of mistake mentioned in the Ready-Peer group were: 'agreement (9 students), 'intonation' (3 students), 'how to answer' (2 students), and 'misuse of interrogatives in the main clause' (1 student). The two types reported in the Ready-Self group were: 'intonation' (2 students) and 'how to answer' (1 student), and none of the students reported successful self-corrections involving 'agreement' and 'use of interrogatives in the main clause'. What are the reasons for the absence of reports on these two aspects?

The absence of reports in the Ready-Self on 'use of interrogatives in the main clause' can be explained

by their relatively high-level of English proficiency. As noted earlier, the students in these two Ready groups had reached stage 5 or 6 on the pretest, and it is conceivable that they rarely made mistakes concerning 'use of interrogatives' in tag questions. In fact, we have seen that in the Ready-Peer the number of reported self-corrections made *with* assistance involving such mistakes was small (1 student), and that in either Ready group there were no reports on self-corrections made *without* assistance concerning such mistakes.

For the absence of reports on the self-corrections involving 'agreement' in the Ready-Self, I should first draw the reader's attention to the finding that, in the Ready-Peer, the reported self-corrections concerning 'agreement' were the most frequently mentioned among the 20 students (9 students=45%). The usage of 'agreement' was focused on in the mini-lecture and the students were aware that the application of *Q-tag* was the target of the two conversation tasks, since almost all questions they were required to produce were tag questions. So, what was the reason for the absence of reports on self-corrections among the 11 students who had watched the self-correction version? It seems reasonable to speculate that the students in both groups made a certain number of mistakes in realising 'agreement' in reconstructing the 10 tag questions in the dictogloss and in 18 tag questions produced in the following information-gap task (*Memory Game*). In fact, we have seen that a similar proportion of students in each Ready group reported the self-corrections made *without* assistance on their mistakes in realising 'agreement' (4 Ready-Peer students=20%, 2 Ready-Self students=18%). These observations suggest that the students who had watched the self-correction version tended to fail in the following three elements of the conversation tasks: (a) noting mistakes in 'agreement' both in their own input and in their partner's, (b) correcting mistakes when assistance was provided, and/or (c) providing the partner with the necessary assistance. These results suggest that the self-correction version of modelling was not as effective as the other in promoting noticing, the resultant self-corrections, and peer assistance in producing target questions.

I should also remind the reader that we have seen in the analysis of the reported benefit of the modelling video that six of 20 students (30%) in the Ready-Peer wrote that the modelling was useful for learning 'the target rule', while none of the 11 students in the Ready-Self did so. I have argued that the peer-assistance version was more effective than the other in making salient the seeded input in the

modelling, and that this eventually made the difference in the proportion reporting learning 'the usage of target rule' as a benefit of watching the modelling. This argument has now been supported by the findings that the successful self-corrections concerning 'agreement' were reported only by those who had watched the peer-assistance version. This suggests that what the students learnt through watching the peer-assistance modelling (i.e. how to assist the partner and the usage of target rule discussed in the process of self-corrections assisted by the partner) had positive effects on the promotion both of noticing the partner's mistakes in realising 'agreement' in tag questions, and of peer assistance to help the partner spot those mistakes and make self-corrections in the subsequent conversation tasks.

These findings give support to my claim that the peer-assistance version was more effective in promoting not just peer-assistance but also noticing and the resultant self-corrections assisted by the partner; but I should again point out uncertainties about using the students' self-reports alone for the analysis. Indeed, we have seen that the proportion reporting successful self-corrections involving 'the target rule', supported by the partner, was larger in the Ready-Peer than in the Ready-Self; however, we should note that these results do not necessarily mean either that the students were able to correct all mistakes in the area mentioned in their reports, or that they made successful self-corrections as reported in the questionnaire. Learners can make wrong corrections, especially when their attention is focused on *meaning*. They can also fail to notice their wrong corrections even when paying attention to *form*. In fact, one Ready student who reported the successful self-correction wrote that he was not able to remember how he had corrected his mistake while filling in the questionnaire, although he was sure that his self-correction had involved 'agreement' in the tag question.

As mentioned earlier, the analysis of the classroom conversation would require further transcribing and is beyond the scope of the present study, but in the next part, to support my discussion, I will quote a sample of classroom interaction in order to give concrete examples that the reported useful peer assistance did take place and that it led the students to successful self-corrections.

8.2.1.3 Evidence of successful self-corrections supported by the partner

I will now quote two classroom conversational episodes which show the process of self-corrections made *with* assistance by two pairs of students (Hide and Mat, and Junko and Fuku) in the second dictogloss after watching the peer-assistance version of modelling video.

Hide and Mat

Hide and Mat, both at the university, were assigned to stage 6 on the pretest, but Mat showed a gap for *Q-tag*, while Hide had acquired this rule. In the questionnaire, Hide reported his successful self-correction involving 'agreement' assisted by Mat and wrote an example, as in the following:

'positive-positive (is wearing …., is he?)'

In this example, Hide meant that his mistake(s) concerned the use of the positive form of the auxiliary 'is' in both the main clause and the tag. I looked for where it had taken place in the class recording and found exactly the same mistake in his utterance in the second dictogloss (see below).

1 Hide: Number 8. Joe is wearing glasses, . is he? ↑
2 Mat: … You said 'Joe is wearing glasses, is he? ↑ '
3 Hide: No, Joe isn't.
4 Mat: Joe is wearing glasses, is he? ↑ . is wearing glasses, is he? ↑
5 Hide: … [Laughs helplessly]
6 Mat: You said, Joe is wearing glasses, is he? ↑
7 Hide: Uh, sorry! [laughs] isn't he? ↑
8 Mat: Yeah, [laughs] rising intonation. It's sure.
9 Hide: Okay. [Laughter from both]

In this extract, Hide was reconstructing a tag question from his notes taken in the second dictogloss task, and Mat was writing down what Hide was saying on the answer sheet provided (see Appendix 22), while checking (a) whether the statement in the reconstructed tag question was correct and (b) whether the target rule was appropriately applied, as encouraged by the teacher before starting the task.

As reported in the questionnaire, Hide used the positive form of auxiliary 'is' in both the main clause and

the question tag (Line 1). On the tape, what the students had actually heard was 'Joe is wearing glasses, isn't he?' and Hide should have used the negative form of auxiliary in the tag. After a few seconds' silence, Mat repeated Hide's utterance with some emphasis on the auxiliary 'is' in the main clause (in Line 2). However, it seems that Hide misunderstood Mat's signal and uttered 'No, Joe isn't' (in Line 3), since he thought that the statement he had heard on the tape should have been 'Joe isn't wearing glasses'.

Mat seemed to have noticed that Hide had got confused, but did not directly point out what was wrong and repeated his original tag question twice (in Line 4). Hide fell silent for a few seconds and then laughed helplessly (in Line 5). At this point, it seems that Hide had still not realised he had made a mistake in agreement and was not sure which part of his tag question was wrong . Mat repeated Hide's original utterance (in Line 6), and immediately after that, Hide noticed his mistake and uttered 'Sorry!' and the correct question tag 'isn't he?' (in Line 7). Hide's laughter after noticing and his cheerful tone of voice in his short utterance 'Sorry! isn't he?' shows that Hide was pleased to have detected the mistake and made the self-correction. Mat also seemed to have been relieved that his partner had noticed the mistake and said 'Yeah', followed by his laughter. Mat then added 'It's sure' meaning 'definitely' to confirm the uncertainty of the speaker's statement on the tape, as the students were required to answer it by ticking either 'sure' or 'not sure' printed on the answer sheet (Line 8). Both Hide and Mat laughed in Line 9 and moved on to the next task question.

This is a good example of peer assistance and evidence that effective self-correction did actually occur *with* the peer assistance, as reported in the questionnaire. Throughout the interaction, Mat was very patient, repeating his partner's utterance to give Hide enough time to *notice* the mistake in his output. In the modelling video, they had watched the model dyad repeating their partner's erroneous utterance to signal a mistake in agreement, as Mat did in the subsequent task. In the video, the model playing the writer's role went on to give a further hint that her partner's mistake had involved agreement, as in 'Umm.. Your intonation was correct, but you said isn't-isn't, negative-negative'. In contrast, Mat did not point out the cause of Hide's mistake and patiently repeated the partner's original utterance and waited until Hide came to notice the mistake for himself. Thus, the way Mat offered Hide assistance was indirect. It is conceivable that Mat believed his partner was able to notice his mistake without any

direct indication of its cause. From their cheerful tone of voice and laughter after the successful self-correction, we can see that they were content with both the outcome (i.e. Hide's noticing and self-correction) and the way it was achieved (i.e. the way peer assistance was given by Mat and the way the assistance was appreciated by Hide). It is worth pointing out that this is how Vygotskians hope corrective feedback takes place (Aljaafreh & Lantolf, 1994). In Mat and Hide's case, there was not a novice-expert relationship, although only Hide showed a gap for *Q-tag*. Their English proficiency was comparable and the one able to spare resources to pay attention more to *form* in the joint activity (i.e. Mat) successfully estimated the appropriate level of assistance and helped his partner perform the given task at his potential level of ability, that is, within Hide's zone of proximal development (ZPD).

I will quote one more dialogue from the classroom data which also constitutes good evidence that the reported successful self-correction did take place in the classroom *with* peer assistance.

Junko and Fuku

Junko and Fuku, both at the university, were assigned to stage 6 on the pretest, but Fuku showed a gap for *Q-tag*, while Junko had acquired this rule. Junko wrote in the questionnaire that her partner's assistance had helped her notice mistakes not only in the use of 'intonation' in tag questions but in 'agreement'. I listened to their recorded conversations in the second dictogloss task and found evidence to support her report, as shown below.

1 Junko: Sarah has small nose, does he? ↑ does she? ↑
2 Fuku: (Err does she? ↑ is .)
3 Junko: Oh, negative-negative.
4 Fuku: (No, no.)
5 Junko: Uh, positive-positive.
6 Fuku: (So. Maybe. So …)
7 Junko: Sorry. Sarah has small nose, doesn't she? ↑
8 Fuku: (No, err . the speaker was sure.)
9 Junko: doesn't she? ↓
10 Fuku: (doesn't she ↓ … Okay.)

As Junko reported, the tag question she reconstructed from her notes contained two kinds of mistake (in Line 1). What the original speaker of the question on the tape asked was 'Sarah has a small nose,

doesn't she?' with falling intonation, and Junko should have used the negative form of auxiliary in the tag. Junko also failed to catch the intonation in the preceding listening and used rising intonation.

Fuku repeated what Junko had said (in Line 2) and this led Junko to think she had made a mistake in agreement. Then, Junko said 'Oh, negative-negative' (in Line 3), which was wrong noticing. Fuku replied 'No, no' (in Line 4) because Junko had used the positive form in both the main clause and the tag, and then Junko noticed what she had said was wrong, saying 'Uh, positive-positive' (in Line 5). Fuku invited Junko to correct her mistake by saying 'So...' (in Line 6) and Junko successfully realised agreement in the tag question (in Line 7). Fuku further pointed out that Junko's 'intonation' was not correct, giving Junko a hint as in the following: 'No, err . the speaker was sure' (in Line 8). This made Junko realise she should have used falling intonation to match the certainty of the original speaker and helped her correct the mistake (in Line 9), which was acknowledged by Fuku, as in 'doesn't she ↓ …Okay' (in Line 10).

In the preceding mini-lecture on *Q-tag* usage, the students had heard how intonation of the tag could convey the certainty of the statement in tag questions, and this extract illustrates how the students actually assisted each other in forming tag questions in that manner. In the above-cited conversation, we can see that Fuku's assistance was more direct than that seen in the other pair. When her partner failed to correct the intonation in the tag (in Line 7), she told Junko that the speaker of the tag question on the tape was certain of her statement, as in 'No, err . the speaker was sure', and hinted that falling intonation had been used by the speaker.

As I noted earlier, the students' self-reports represent indirect evidence, but the above-cited two extracts from the classroom conversations did show that useful peer assistance reported in the questionnaire actually took place in the class and led the students to make successful self-corrections, as they reported. Given that about half of the 20 Ready students who had watched the peer-assistance version reported self-corrections made *with* peer assistance involving 'agreement', while none of the 11 Ready students who had watched the self-correction version did, it seems reasonable to say that the type of modelling video affected students' classroom behaviour. It is also possible to say that the peer-assistance version had an advantage over the other in terms of stimulating them to provide peer assistance to help the

partner make self-corrections.

8.2.1.4 Classroom data showing the limitations of using self-reports alone

In the previous section, although I have shown two extracts which exemplify successful self-corrections supported by the partner, I have also found evidence while examining the classroom interaction data that a student who was able to self-correct an 'agreement' mistake in one tag question *with* assistance from the partner in the dictogloss task nevertheless failed to notice the same mistake in the following information gap task.

Koara and Mama

Koara and Mama, both at the university, were assigned to stage 6 on the pretest. Neither showed a gap for *Q-tag*, but Koara reported in the questionnaire that she had made mistakes in agreement in tag questions (positive-negative/negative-positive) and had received useful assistance from Mama in correcting them. While I have found evidence to support her report in the recording of the second dictogloss task, in which similar assistance observed between Hide and Mat took place, I have also seen that she did not correct another two mistakes involving agreement in the subsequent information-gap task called *Memory Game*.

Before presenting the extract, I will briefly explain the procedures for the *Memory Game*. Each pair of students was provided with a sheet of paper on which four portraits were printed in colour (see Appendix 13). They were asked to memorise each person's characteristics on the portrait, so that they could use their memory in the subsequent speaking task. The students were given one minute to do so, then told to turn over the portrait sheet. Then, the students in each pair were given one of two answer sheets, (A) or (B), on which they found nine blanks to describe the facial characteristics of the four portraits (Appendix 13). On each answer sheet, nine different places were left blank, and the students individually were asked to fill them in from memory. After every student filled in the blanks, they were instructed to ask their partner questions by using the target rule (*Q-tag*) to check whether their memory of the characteristics of each portrait was correct. The other student in the pair needed to listen

to the questions very carefully because they did not know which characteristic of which portrait their partner would ask about. They also did not have notes on those characteristics and needed to answer from memory. Thus, this information-gap task was more demanding than the dictogloss task in which both students in a pair took notes while listening to the same tag questions to be reconstructed later.

I will now cite Koara and Mama's conversation from the Memory Game. In the following extract, Koara was asking questions to find out if her memory of the nine characteristics of the four portraits was correct. Mama was answering the questions – her answers were put in round brackets.

1. Mario has . black hair, doesn't he? ↓ (Yes, he does.)

2.Mario has a /mustache/, . doesn't he? ↓ (I don't think so.) No? no? Are you sure? (Mmm) In fact, I don't remember. (I don't remember neither) Mario has, I mean, Mario has a mustache, doesn't he? ↑ laughter both (I don't think so). Next one (okay) okay

3.Mario Mario is not wearing glasses, is he? ↓ (No, he isn't.)

4. Next Herman. He is bald, isn't he? ↓ (Yes, he is.)

5. Herman has . Herman doesn't have beard, does he? ↑ (No, he doesn't.)

6.*Herman is not wearing a ribbon, . does he? ↓ (No he doesn't.) No, no, of course not.

7. *Sally is wearing a ribbon, . doesn't she? ↓ (Yes, he does) she (she) she does. (okay)

8. Paul has white hair, doesn't he? ↓ (Yes, he does)

9.He has a small nose, doesn't he? ↓ (Erm… I don't think so). No? (I don't think so.) You mean he has a big nose? ↓ (I think so) Are you sure? (No..) I'm sure he has a small nose.

In the first five tag questions, Koara produced correct tag questions with agreement. We can see from their interaction that both students were concentrating on meaning. For instance, in the second example Koara thought Mario had a moustache and asked 'Mario has a /moustache/, . doesn't he?' ↓ , but Mama did not agree and said 'I don't think so'. Then, Koara asked Mama 'No? no? Are you sure?'. Both students then admitted that they did not have a clear memory of Mario's moustache and moved on the next question.

Koara's mistakes in agreement occurred in the sixth and seventh examples. In the sixth example the auxiliary in the tag did not agree with that used in the main clause, as in 'Herman is not wearing a ribbon,

does he?'. Both students seemed not to have noticed this mistake, and Mama replied 'No he doesn't', followed by Koara's utterance 'No, no, of course not' to confirm that her memory was correct. We can see that both students were focusing on meaning – whether the statement of the tag question was correct or not. In the next question, although Koara made a correct statement in the main clause ('Sally is wearing a ribbon'), she used an incorrect auxiliary in the tag ('does he?'). This mistake was also left uncorrected.

Both students had acquired *Q-tag* from the outset and looked confident while carrying out the tasks. In fact, Koara made an additional comment in the questionnaire that she had never thought she would make agreement mistakes in tag questions and she was surprised when she noticed her mistake through her partner's assistance in the dictogloss task. There is a possibility that Mama had noticed the mistakes in the conversations quoted above but did not raise them in the task, as the task goal was to get the answers right. However, there is good reason to think that Mama was too busy replying to the questions and did not notice the mistakes.

There is a difference between English and Japanese in answering questions with a tag, and Japanese learners tend to have difficulties in replying 'yes' or 'no', especially when the negative form of verb/auxiliary is used in their statement. For instance, to tag questions like 'Herman isn't wearing a hat, is he?', Japanese learners are inclined to reply 'Yes, he isn't' as they do in Japanese, since they prefer to acknowledge that the statement was correct by saying 'yes' to mean 'Yes, (as you've said) he isn't wearing a hat'. It is conceivable that Mama directed her attention both to the choice of correct word (i.e. 'yes' or 'no') in her replies and to the content of statement in her partner's tag questions, and that this burden distracted her from noticing her partner's mistakes.

The analysis of the extracts from the classroom conversations have shown both the possibilities and limitations of learners' ability to notice mistakes in output, both their own and their partners'. In the case of Koara and Mama, they were able to notice the mistakes in the dictogloss task but not in the Memory Game which was cognitively more demanding.

Added to this, while looking for evidence for the self-correction reported by Junko, quoted above, I

found direct comments made by Junko and her partner, Fuku, on their difficulty in paying attention to their partner's output in providing assistance in the Memory Game. The extract below was recorded after they were given the answer sheet for the task to check whether their memory of the portrait was correct. We can see from Fuku's confession that she had reflected on her performance after completing the task. Junko's utterances are in the round brackets.

> **Fuku:**
> '*Zenzen [=at all]*, tag question . I don't didn't hear your tag question . right or wrong. I didn't hear (I understand.) I didn't take care *jyanai [=no]* . I didn't pay attention to your tag questions . whether they were right or not . (Me, too) Just memo note.'

These are examples of 'confessions' in private chat. I should remind the reader that in the preceding dictogloss task, which was less demanding than the *Memory Game*, Fuku's assistance helped Junko to notice her mistakes in agreement in the tag question, and in intonation to express the certainty of statement. Fuku's comments show that she knew she was supposed to check whether her partner's use of *Q-tag* was appropriate. It seems, however, that she had been too preoccupied by her role to reply to her partner's tag questions from memory. Junko acknowledged Fuku's comment as in 'I understand' and 'Me, too', admitting that it was also difficult for her to pay attention to the use of the target rule in Fuku's questions. In fact, I found two mistakes in agreement in tag questions produced by each student in the Memory Game, as shown below.

> Fuku: 'Herman doesn't have beard, doesn't he? (.. Herman doesn't have beard.) Yes, (Yes.) okay.'
>
> Junko: 'Herman has a red hair . has red hair, . does he? ↑ (Yes, he has. read hair.)'

The first example was produced by Fuku and the negative form of auxiliary 'does' was used in both the main clause and the tag. Junko repeated the main clause, but the mistake was left uncorrected. The second example was produced by Junko who should have used the negative form of auxiliary in the tag, but used its positive form. Her partner repeated the main clause in the tag question and acknowledged that Fuku's statement was correct; however, the mistake in agreement was not discussed between the two and was left uncorrected.

These instances suggest the importance of teachers' choice of tasks and of sequencing them to promote noticing. In this study, I used the form of dictogloss to help learners become aware of their L2 production and we have seen evidence in the students' self-reports and classroom interaction that it served its purpose and offered a productive route to noticing. I hoped that their noticing in the dictogloss task would encourage them to carry on paying attention to mistakes in both their own output and in their partners' in the subsequent information-gap task. The findings, however, show the limitations of the students' ability to notice mistakes in using the target rule in the cognitively demanding task. This suggests a need to provide a post-task activity, as a route to noticing, such as a transcribing task from the classroom recordings, which provides learners with opportunities to reflect on language use. (See Clennell, 1999; Lynch, 2001, 2005, for manageable and effective ways of using transcribing to encourage learners to notice forms in L2 output and to improve accuracy in their use.)

Having shown the results of my analysis of the reported successful self-corrections made *with* or *without* partner assistance and some supporting evidence that the reported useful peer assistance in the self-reports did take place in the classroom, I will now turn to examine the self-reports made by the two groups of Unready students in the same questionnaire.

8.2.2 The two Unready groups

The research question formulated to examine the effect of the two types of modelling video was:

Q8. Are there any differences in the noticing (i.e. reported successful self-corrections) of Unready students in the subsequent tasks depending on the type of modelling video they have watched?

8.2.2.1 Number of reported successful self-corrections

Table 8.6 below summarises the number of Unready students who reported their own self-corrections *with* and *without* assistance from their partner. I will first examine the reports on successful self-corrections made without peer assistance.

TABLE 8.6 Number of students reporting self-corrections (Unready)

	Number of students	
	Target rule (Q-tag)	Non-target L2 items
***Without* assistance**		
Unready-Peer group (N=9)	4 (44%)	1 (11%)
Unready-Self group (N=7)	2 (29%)	2 (29%)
***With* assistance**		
Unready-Peer group (N=9)	1 (11%)	1 (11%)
Unready-Self group (N=7)	1 (14%)	4 (57%)

Self-corrections *without* assistance

In both Unready groups, successful self-corrections made *without* partner assistance were reported. For those involving 'the target rule', four of the nine students (44%) in the Unready-Peer made a report., while in the Unready-Self two of the seven (29%) did so. The content of the reports is summarised in

Table 8.7 below.

TABLE 8.7 Number of self-reports on successful self-corrections in the two Unready groups (Week 1)

	***Without* assistance**		***With* assistance**	
	Unready-Peer (N=9)	Unready-Self (N=7)	Unready-Peer (N=9)	Unready-Self (N=7)
Target Q-form (*Q-tag*)				
● Agreement (negative-positive/ positive-negative)	3 (33%)	2 (29%)		1 (14%)
● Intonation to express certainty of the statement	2 (22%)	1 (14%)	1 (11%)	
How to answer tag questions				
Use of interrogative in the main clause				
Non-target L2 items	1 (11%)	2 (29%)	1 (11%)	4 (57%)

Note. The items marked by ● were dealt with in both versions of modelling video as mistakes.

The reported successful self-corrections in the Ready-Peer involved 'agreement' (3 students) and 'intonation to express certainty of the stagement' (2 students) in tag questions, which were focused on in the mini-lecture. In the Unready-Self, two of the seven students (29%) reported successful self-corrections in realising 'the 'agreement' and one student made a report on a successful correction in using 'intonation to express certainty of the statement. These results show that a larger proportion of students who had watched the peer-assistance version had the experience of correcting mistakes in applying the target rule, but for the content of the self-corrections, the students in each Unready group mentioned the two focused usages of tag questions (i.e. 'agreement' and 'intonation').

For the reported successful self-corrections involving 'the non-target L2 items' in the two Ready groups, the number of students reporting this was only one of nine (11%) in the Unready-Peer and two of seven (29%) in the Unready-Self.

Self-corrections *with* assistance

I will now examine the self-reports on successful self-corrections made *with* assistance. For those involving 'the target rule' (*Q-tag*), only one student in each Unready group reported self-correction (see Table 8.7 on the previous page). One correction reported in the Unready-Peer involved 'use of intonation to express certainty in tag questions' and one reported in the Unready-Self concerned 'agreement (positive/negative) in tag questions'. The small number of self-reports in both groups may be explained by their unreadiness to learn the target rule. Most Unready students worked with a partner who was also unready to learn the target rule, and it is conceivable that they had difficulties in both applying the rule and providing assistance while carrying out the tasks. There is a possibility that they received useful assistance but were unable to use it to make self-corrections. It is also possible that they made successful self-corrections *with* assistance but did not notice that.

As for successful self-corrections concerning 'the non-target L2 items', the number of students reporting a positive experience was only one of the nine (11%) in the Unready-Peer and it involved ''the misuse of verbs'. On the other hand, in the Unready-Self the proportion writing a report which helped make corrections in using 'the non-target L2 items' was bigger (4 of 7=57%) and their reports involved 'misuse of verbs' and 'meaning of words'. One possible reason can be suggested for the difference between the two Unready groups. As we have seen in Section 8.1.2.2.1, the students who reported that the modelling was useful for learning 'how to assist the partner' were without exception those who had watched the peer-assistance version of modelling, and it is conceivable that they were more motivated to offer assistance than those who had watched the other version. It is possible that they paid much more attention to the partner's mistakes in using the target rule, as done by the models on the video, but their effort did not result in successful corrections and few reports were made in the questionnaire. Again, analysis of classroom data is necessary to see what really happened in the class, but it is beyond the scope of this study.

8.2.3 Summary for Section 8.2

We have seen that the proportion reporting on successful self-corrections involving 'the target Q-form' made *with* partner assistance was larger in the Ready-Peer (12 of 20=60%) than in the Ready-Self (3 of 11=27%) and I concluded that the peer-assistance modelling video was more effective for the Ready students in promoting self-corrections. Many of these reports involved 'agreement' and 'intonation' in tag questions, which had been focused on in the treatment. As for the reported successful self-corrections made *without* assistance in producing tag questions, no significant difference was observed between the two groups (8 of 20 Ready-Peer students=40%, against 4 of 11 Ready-Self=36%).

In addition, I provided concrete examples of peer assistance and self-correction of mistakes in realising 'agreement' and in using 'intonation' in tag questions, as evidence that the reported successful self-corrections assisted by the partner had actually occurred in the classroom. I have also discussed the evidence from the classroom data that the reported self-corrections involving 'agreement' did not take place every time the students made mistakes in realising 'agreement, and I have recalled the limitations of using the self-reports alone for the analysis of learners' successful self-corrections. We have also seen that the students' noticing was influenced by the cognitive demands of the task. Although the number of examined extracts from the classroom data is limited, we have seen that in the more demanding task (Memory Game) the students were unable to notice agreement mistakes in tag questions, which they had been able to notice and correct in the less demanding task (Dictogloss question reconstruction task).

As for the Unready students, the number reporting on the successful self-corrections made *with* partner assistance in using 'the target Q-form' was only one in each group and I have argued that the lack of report was due to their unreadiness to learn the rule. On the other hand, in these two Unready groups, the proportion reporting successful self-corrections made *without* partner assistance in using 'the target Q-form' (4 of 9 Unready-Peer students=44%, 2 of 7 Unready-Self= 29%) was comparable to that found in the two Ready groups (8 of 20 Ready-Peer students=40%, 4 of 11 Ready-Self=36%). This suggests

that, although the majority of Unready students did not recall a positive experience in making self-corrections made *with* partner assistance in using the target rule, the six Unready students mentioned above had a positive experience in producing tag questions *without* peer assistance. I have not provided supporting evidence from the classroom recordings and should be cautious in drawing a firm conclusion here; however, these results show that more than 30 percent of students in the four groups noticed their successful corrections made *without* partner assistance in using the target rule.

As for the reported successful self-corrections with or without partner assistance in using 'the non-target L2 items', the proportion writing a report was relatively small in the four conditions (11-29%). The only exception was found in the Unready-Self group, where more than half Unready-Self students (4 of 7=57%) reported successful self-corrections made *with* assistance in using the 'non-target L2 items. I have discussed the likelihood that their unreadiness prevented them from offering assistance in producing the target rule, although they cooperated with each other and managed to help the partner make self-corrections using 'the non-target L2 items' which they had acquired and/or they were ready to learn.

Bearing in mind the limitations of basing my analysis mainly on self-reports from a small sample, it is possible to conclude that, as long as the target forms seeded in the modelling were at the next stage of students' L2 development, the peer-assistance version was more effective in promoting self-corrections made *with* assistance in the subsequent conversation tasks. This should encourage teachers when preparing modelling to take into account the students' readiness for the L2 grammar item.

I will now turn to the students' reports on their positive changes in using the target rules beyond the treatment sessions.

8.3 Positive experience involving the target rules beyond the classroom

In the first treatment session in each class at each institution, I explained the aim of the course to the students (i.e. to ask questions using various Q-forms, especially the target ones, in communication) and I emphasised the importance of *noticing* in L2 learning in attaining that goal. I told them that their effort to pay attention to their own output and their partner's would help them notice the gap between what they thought they had known and/or heard about the target rules in the mini-lesson and what they could actually do with them, and that it would eventually help them correct their mistakes and acquire the rules.

I also said that in Japan people tend not to hear questions, such as tag questions, in which the target rules are applied, especially in the classroom, but that the target rules are often used by native English speakers in daily conversations to express ideas and feelings which would enrich the relationship. I added that, once they experienced noticing the gap through the treatment, they would also become able to notice examples of the target rules in someone else's speech in conversations, in person, and in films or TV programmes. I added that they would be surprised that they had not noticed them before.

They looked half-convinced. I told them that this was my own experience as a learner of English, and that of my Japanese friends back in Edinburgh. They also seemed doubtful about the chance to experience such noticing in Japan, where English is not used for daily communication. It is natural for learners to have doubts like these and I was interested to examine their self-reports of positive experience involving the target rules both in the class and beyond.

In the previous two sections of this chapter, I have analysed the students' experiences in the treatment, so I will now examine their reports on positive experience beyond the treatment sessions. I collected the data from the questionnaire given immediately after the follow-up test. I asked them to reflect on their positive experiences outside the class from two angles: (a) their use of target rules and (b) noticing the use of target rules in someone else's speech (see Appendix 82 for the questionnaire). The students had the winter holidays after the last treatment session, and by the time they filled in the questionnaire six or seven weeks had passed, which means that their reported experiences outside the classroom took place

within three months of the first treatment session.

8.3.1 Findings and discussion

I formulated the following two research questions to examine the students' positive experience involving the target rules beyond the classroom:

Q9. Do the students notice any positive changes in their use of the target rules beyond the treatment sessions?

Q10. If so, what kinds of changes do they notice?

In analysing their self-reports, I first counted the number of students who made a report, then examined the content of their experiences.

8.3.1.1Number of students who reported a positive experience

The number of students reporting positive experience is summarised in Table 8.8 below, in relation to their developmental stage on the pretest.

TABLE 8.8 Number of students reporting positive changes involving the target rules

Developmental stage on the pretest		Numberof students who reported positive changes	
Stage 6	(N=22)	12	(55%)
Stage 5	(N=9)	8	(89%)
Stage 4	(N=4)	0	(0%)
Stage 3	(N=11)	1	(9%)
Stage 2	(N=0)	-	(-)
Stage 1	(N=1)	0	(0%)
Unknown	(N=1)	1	
Total	(N=48)	22	(46%)

Of the 48 students who had received the treatment, 22 students (46%) made reports. The reports given by 17 of the 22 students involved one of the four target rules (i.e. *Cancel-inversion, Q-tag, Negative-Aux/Do-2nd, Y/N-negative*) and the other reports made by five students concerned more than one target rule.

These students comprised eight of nine at stage 5 (89%), 12 of 22 at stage 6 (55%), one of 11 at stage 3 (9%), and one student whose developmental stage was unknown[37]. These results show that positive experience involving the target rules after receiving the instruction was reported mainly by those at stage 5 or 6. The small proportion of students found among those at stage 4 or below can be explained by their non-acquisition of the stage 6 Q-forms through the instruction. In fact, as we have seen in Chapter 7, most Unready students did not acquire any target rules, and it is not surprising that they did not have positive experiences in using the target rules and in noticing the use of target rules in someone else's speech outside the classroom[38].

It should be noted that the proportion of students reporting positive experience was larger among those at stage 5 than among those at stage 6. Two reasons can be given for this. First, it is conceivable that those who had been taught the Q-forms at the next stage of their development (i.e. stage 6 rules) were more motivated to try out the newly learnt Q-forms than those who had been able to use those rules. As shown in Chapter 6, the instruction focusing mainly on the stage 6 rules was most beneficial for the students at stage 5 in terms of developmental stage advancement, and it is conceivable that they wanted to report their gains in the questionnaire. Second, it is reasonable to think that before taking the treatment, many stage 6 students had experienced what the stage 5 students reported as a positive experience and that was the reason for the smaller proportion of students who made the report in this questionnaire.

I also examined the number of students who made a report on each target rule in relation to their acquisition of each Q-form on the pretest, and its results are summarised in Table 8.9 on the next page. For *Cancel-inversion*, reports were made by 13 of the 48 students (27%), including three who had acquired this rule and 10 who had not. For *Q-tag*, a total of 13 students made a report, including five who had acquired the rule and eight who had not. These figures show that about 30 percent of students

who had received the instruction reported positive experience involving these target rules beyond the treatment sessions. For each rule, the proportion of students who made the reports was larger among those who had not acquired the rule than among those who had.

In contrast, very few reports were made on the other two target rules. Only one student mentioned *Negative-Aux/Do-2nd* and no student wrote a report on *Y/N-Negative*. One possible reason is the lack of opportunity to use these rules in daily life in Japan. For instance, although in the treatment the students used *Y/N-negative*, as in 'Haven't you done the shopping yet?', to show that speaker was surprised that something had not happened, the opportunity to apply the rule in this way seems rare outside the classroom. Another possible reason is that they had a smaller amount of practice in using these two rules in the treatment than the other two. As noted earlier in the previous three chapters, *Negative-Aux/Do-2nd* and *Y/N-inversion* were taught together in the last session, while *Cancel-inversion* and *Q-tag* were taught separately in the first two classes. Moreover, they had further practice in using *Cancel-inversion* and *Q-tag* in the third class.

TABLE 8.9 Number of students reporting positive changes after the instruction

Target Q-form	Acuisition of the rule on the pretest		
	Acquired	Not acquired	Total
Cancel-inversion	3	10	13
Q-tag	5	8	13
Negative-Aux/Do-2nd	0	1	1
Y/N-negative	0	0	0

8.3.1.2 Content of the positive changes

I will now examine the content of reports in detail. The reported positive experiences can be categorised into two: (a) noticing the use of target rules in someone else's speech and (b) noticing improvement in producing questions realising the target rule. I will show the results of my analysis for

each category by quoting the reports. It should be noted that all reports were written in Japanese with or without example sentences in English, so I have translated the cited comments into English. Where necessary, I added some words to the original reports to make the meaning of comments clearer. I also added some contextual information for better understanding of the reports and put it in round brackets.

I should also note that reports made by those who had acquired the rule on the pretest are marked with a '+'. Reports marked with a '-' show that they were given by students who had not acquired the rule on the pretest. If the '-' is followed by 'A', as in '- A', it means that the acquisition of the rule occurred after the instruction (i.e. the production of two productive usages of the rule on the two posttests). If it is not followed by 'A', as in '-', it means that the rule was not acquired after the instruction.

8.3.1.2.1 Noticing the use of target rules in someone else's speech

There were five students who, after receiving the instruction, reported the experience of noticing the use of tag questions in someone else's output outside the classroom. For example, one student (Junko) reported that she had noticed the tag questions used in the modelling video when she was attending a course given by a native English speaker, as in the following.

+ After doing the tasks in the lessons, I began to notice the structure *Q-tag* in the (model) conversations given in other (English courses) and in talks given by an English lecturer.
by Junko [Maintained stage 6]

The other four students also wrote about experience of noticing the use of tag questions in films as a positive incident, as follows.

- A After taking the lessons, I began to notice the use of tag questions while watching films and in English learning materials.
by Hana [Maintained stage 6]

- A I began to notice the use of tag questions more often while watching films on video.
by Nae [Sustained stage gain from 5 to 6]

+ I think I began to be able to notice tag questions while watching films, although not all the time.
by Hide [Lost a stage from 6 to 5].

- While watching the film, Dracula, on DVD yesterday, I was startled every time someone produced a tag question. It seems my body has become *Q-tag*-sensitive. by Shin [Lost two stages from 5 to 3]

In these reports we can see the students' pleasure in their new experiences. For learners in the EFL context, watching films is one of the main opportunities to listen to English, so noticing L2 items which they used not to notice in conversations between native English speakers is the moment when they are most aware of improvement. For example, Shin described his delight jokingly in his report, expressing that he was pleased to have become able to notice tag questions used in a film – in his words, to have become '*Q-tag*-sensitive'. Although Shin was considered not to have acquired this rule, he started producing many examples of *Q-tag* on the posttests (three examples, of which two were productive examples on the posttest, and six examples with person and number agreement on the follow-up test, of which one five were invalidated because of his use of same auxiliary verb). This suggests that reconstruction of his interlanguage had been under way through trial and error.

8.3.1.2.2 Perceived improvement in noticing and producing the target question

Reports involving perceived improvement in noticing and producing the target rule question were made by 22 students. The reported positive experiences took place on various occasions: in the treatment sessions, on the posttests after the instruction, beyond the treatment and tests given in this study, such as in other English courses given by a native English speaker, while watching films on video at home, and while working part-time at cram school as an English teacher and at a fast-food restaurant.

On the posttests

There were 13 students who reported improvement in producing questions with the target rule on the posttests. Five students' reports concerned *Cancel-inversion*, one student's report involved *Negative-Aux/Do-2*nd, and seven students' reports concerned *Q-tag*. For example, the following three reports involved improved performance in applying *Cancel-inversion* on the posttests (Taka, Jo, Tana). All three students gained a stage from 5 to 6 and these reports and the acquisition of this rule occurred after the instruction. Their reports clearly show that they were aware of their gains and of their

improvement in applying this rule in the conversation tasks given on the posttests.

- On the test I tried to use polite, indirect questions by using 'Could you tell me..?' and other newly learnt question forms in Task 1 (Guided Role-play) and in Task 2 (Find the Difference).
  by Taka [Stage 5-5-3]

-A After the treatment, I began able to use indirect questions in the speaking tasks on the (post) tests.
  by Jo [Sustained stage gain from 5 to 6]

-A After the treatment, I became able to use the polite question 'Could you tell me..?' in Task 1 (Guided Role-play).
  by Tana [Sustained stage gain from 5 to 6]

In addition, two more students at stage 6 on the pretest (Jin and Tami) reported better performance in using *Cancel-inversion* on the posttests. Jin, whose acquisition of this rule took place after the instruction, wrote that he had consciously made efforts to apply the newly learnt Q-form, and that he was satisfied with his improvement. The other student, Tami, who had acquired this rule before the instruction, also expressed her satisfaction at better posttest performance.

-A I tried to use 'Could you tell me..?' as many times as possible in Task 2 (Find the Difference). I was satisfied when the question 'Could you tell me what the time is?' came out of my mouth smoothly.
  by Jin [Maintained stage 6]

+ I became able to ask questions very well by using 'Could you tell me..? in Task 1 (Guided Role-play).
  by Tami [Stage 6-6-5]

As regards positive experience of producing *Negative-Aux/Do-2nd* questions, only one student (Jo) made a report which involved improved performance on the follow-up test. He gave an example of his successful application of this rule in the report, as in the following.

- On this test I became able to use negative questions and asked 'Why haven't you done yet?'.
  by Jo [Sustained stage gain from 5 to 6]

For positive experience in producing tag questions on the posttests, seven students wrote comments. Two of the seven mentioned their improved performance on the posttests and the other five wrote about

their increased knowledge about *Q-tag*. For instance, in the following two reports, one student (Tana) wrote that she began to use *Q-tag* after the instruction. She had acquired this rule after the instruction and her report shows that she was aware of her gain.

- A On the posttests I began to use *Q-tag* in Task A (Guided Role-play).
  by Tana [Sustained stage gain from 5 to 6]

Maya also realised her improvement on the posttest when she was producing tag questions, but admitted that her success was limited. She could not acquire *Q-tag* through the instruction and remained at stage 3 on the three tests; her report reflected this.

- I learnt how to use tag questions. I can now use the rule a bit, but can't use it appropriately all the time.
  by Maya [Stage 3-4-3]

The other five students' reports concerned their increased knowledge of *Q-tag*, especially their noticing of usefulness in being able to use *Q-tag* in conversations, which enabled them to feel confident in asking questions with *Q-tag*. As noted earlier, in the first treatment session, I asked the students if they had used tag questions in oral communication. Most students said that they hardly ever used tag questions, even in English classes, although they had been taught the usage of *Q-tag* at a senior secondary school and were supposed to have some knowledge of the rule. The following reports represent their state of knowledge of *Q-tag* before the instruction.

- A Until I took the lessons, I didn't know how to use question tag, but I think I can now use it to some extent.
  by Toyo [Stage 5-3-6, showing delayed effects of the treatment]

- I hadn't had a chance to learn how to use tag questions for communication before, and the lessons were very useful for me.
  by Riko [Lost a stage from 6 to 5]

  (After the instruction,) I came to know that the use of tag questions was one of the options for asking questions.
  by Mie [Stage unknown][39]

The following two reports also show that their knowledge of *Q-tag* was limited before the instruction. For example, Seigo admitted that he did not know that intonation in tag questions made a difference in

meaning (i.e. the degree of certainty of the statement). Through the instruction he began to pay attention to intonation and was satisfied with his gain (i.e. the new knowledge about the rule). Shin's report shows that he did not know that tag questions could be used to ask for agreement before the instruction. His report shows that he noticed that tag questions were very useful in carrying out the information gap tasks given in the tests.

+ I didn't know intonation in the end of tag questions made difference in meaning until I took the classes. Since then, I began to pay attention to intonation (when producing tag questions). I'm glad I found out about it.
by Seigo [Maintained stage 6]

- I've realised the usefulness of the use of tag questions in the context where I wanted to confirm something and when I asked for agreement in Task 2 (Spot the Difference). This was really useful expression to carry out the task.
by Shin [Lost two stages from 5 to 3]

These reports show that the instruction helped the students not only to notice the gap in their output but to learn the appropriate usages of *Q-tag*.

In other English courses

The students' positive experience of using the target rule after the instruction was not limited to the treatment sessions and tests given in this study. Five students wrote that they started asking questions by using the target rule in conversation with an English lecturer after receiving instruction in other English courses. For example, two students (Yuta and Nae) reported experience of using *Cancel-inversion* questions, as follows.

+ (On the post tests) I was able to use the polite questions (=*Cancel-inversion* questions). Through the treatment I became able to use this polite way of asking questions and I've used it in other English classes, such as one taught by a native English speaker.
by Yuta [Maintained stage 6]

- A I became more aware of the usage of indirect questions, especially its word order. When I asked questions to my teacher in other course (who was native speaker of English) and asked her to do something for me, I tried to use a polite, indirect questions.
by Nae [Sustained stage gain from 5 to 6]

In these two reports we can see the students' positive attitude towards trying out the newly acquired rule

in communication beyond classroom practice. Both students' acquisition of this rule occurred after the instruction and their reports clearly show that they were aware of the gains, which allowed them to express politeness. We can read their satisfaction between the lines.

The following three reports concern positive experience of using *Q-tag* beyond the treatment sessions and the tests given in this study. Yuta mentioned that he started using tag questions in other English courses and became able to correct his mistakes, as he had been encouraged to do in the instruction. Ken reported his conscious effort to use tag questions in conversations and Tana reported her new experience of using tag questions in email communication, as follows.

\+ After studying *Q-tag* in the lessons, I started using it in other courses. Of course, I became able to notice my mistakes (in using *Q-tag*) and correct them by myself.
by Yuta [Maintained stage 6]

\+ (After the treatment) I tried to use tag questions when I had conversations in English.
by Ken [Stage 6 – stage 5 – stage 6]

-A I also started using tag questions in my email communication.
by Tana-A [Sustained stage gain from 5 to 6]

At part-time jobs

Three students reported experience of using the target rule outside the classroom. For example, two students (Toby and Nobu) wrote that the instruction was helpful in understanding clearly the usage of *Cancel-inversion*, and that this made them feel confident in giving examples of this rule when teaching English part-time at a cram school.

-A I became able to use tag questions and (indirect questions by using) 'Could you tell me..?'. When I taught English at a cram school (after the treatment), I could use them in giving sample sentences to my students
by Toby [Maintained Stage 6]

-A I became able to explain how to use indirect questions when I taught English to lower secondary school students at a cram school between last December and this January. (On the follow-up test) I noticed I became able to variety of question (forms) in Task 1 (Guided Role-play).
by Nobu [Sustained stage gain from 5 to 6]

The first student's (Toby) acquisition of this rule took place after the instruction and his report shows that he was aware of his gain. The second student (Nobu) produced no productive examples of *Cancel-inversion* on the pretest and was considered not to have acquired this rule and his report shows that he had little confidence in explaining its usage until he received the instruction. I remember that Nobu ran up to me in the corridor at the university with a smile on his face about three weeks before the follow-up test to tell me about this experience at the cram school. He also told me that he began to feel funny when he forgot to cancel inversion in the subordinate clause of indirect questions, but was now confident in using this rule, as cited below. We have seen the same comment given by Yuya in the class in Chapter 7. Nobu and Yuya's comments suggest that it takes time for learners to internalise the newly learnt rules since that process can engage them in the struggle to restructure their interlanguage grammar.

- A I felt something was wrong when I inverted the subject and auxiliary (in the subordinate clause) in indirect questions and corrected my mistakes.
by Nobu [Sustained stage gain from 5 to 6]

One student (Hide) also reported his positive experience in using the target rule at a fast-food restaurant as a 'checker'.

- When I was working at MacDonald after the treatment, I had a customer who spoke English. One of the things the customer ordered took a bit long time to prepare. I was pleased that I asked the customer 'Would you mind if you wait for your order?'. Before receiving the lessons, I was not sure about the word order after an 'if' and didn't have confidence (in using this Q-form) . I inverted (the subject and auxiliary) sometimes and didn't invert in some other cases.
by Hide [Lost a stage from 6 to 5]

Hide was assigned to stage 6 on the pretest but showed a gap for this rule, and his report shows that his knowledge about *Cancel-inversion* improved through the instruction and that he felt satisfied when he was able to use the rule in real communication with a native English speaker. He was not considered to have acquired this rule after the treatment, but we can see from this report a sign that the restructuring of his interlanguage was under way.

Noticing the gap in the posttests

There were three students who reported noticing a gap between what they thought they had known about the target rules and what they could actually do with them as a positive experience, and their realisation of the need for practice in using the rules. For example, in the following three reports, the students reported experience involving *Cancel-inversion*. In the first report below, Saya wrote that she noticed through the instruction that there was a gap between what she thought she could do in producing *Cancel-inversion* questions and what she was actually able to do. She also wrote that she had not used *Cancel-inversion* in conversation before and realised the importance of practising the rule to become able to use it in communication.

+ I realised that I was not able to use what I thought I could use, especially those I usually didn't use. It's important to have opportunities to use them through speaking tasks.
by Saya [Maintained stage 6]

Kaji also emphasised the importance of having more practice to use *Cancel-inversion*, as in the following.

-A I can now use polite, indirect questions by using 'Could you..?'. I want to use this as much as I can from now on. But I think that I made mistakes when I paid too much attention to the form on the test. I should get accustomed to applying it.
by Kaji [Sustained stage gain from 5 to 6]

Kaji's acquisition of this rule occurred after the instruction and his comments clearly show his strong motivation to try out this rule, since he realised the value of being able to ask questions in the polite form. He pointed out the need to practise it more, in order to be able to apply it appropriately without paying too much attention to the *form*.

Yan reflected on his use of *Cancel-inversion* in the posttests, referring to the two examples of this rule introduced in the treatment session (i.e. 'Could you tell me....?' and 'Would it be alright if ...?"), as shown below.

+ Through the lessons I've mastered the use of indirect questions by using 'Could you tell me..?' and I think I've achieved good results in Task 1 (Guided Role-play) on the test. In the class I also learnt the expression 'Would it be all right if...?' but didn't use it after

that. This expression was much more unfamiliar to me as an indirect question than the other expression 'Could you tell me..?', and that's why I didn't use it.
by Yan [Maintained stage 6]

Yan admitted that he had not used the latter in the posttests and concluded that his non-use of this expression stemmed from his lack of familiarity. Indeed, the students produced indirect questions using both expressions, but none of them asked questions using 'Would it be alright if....?' in the posttests. One reason is that the examiners did not prompt them to use this expression when eliciting examples of *Cancel-inversion*. Yan's comment suggests a need for seeding occasions in achievement tests to use the variety of expressions taught in instruction since learners tend to use familiar ones, especially under communicative pressure; and unfamiliar ones can be forgotten after the instruction if learners are not given a nudge by the teacher. Testing reinforces learners' learning and Yan's report reminds us, as teachers, that we need to make sure that tasks designed for testing include occasions to use what learners practised in the class.

8.3.1.3 Summary for Section 8.3

I will summarise the findings for the two research questions below.

Q9. Do the students notice any positive changes in their use of the target rules beyond the treatment sessions?

We have seen that about half the students (22 of 48=46%) who had received the treatment reported their positive experiences involving the target rules beyond the treatment sessions, and that many of the positive experiences they reported matched with their achievement in the acquisition of each target rule. However, it has been shown that most reports were given by the higher-level students at stage 5 and stage 6. I have argued that the difference was brought about by their readiness to learn the target rules and pointed out the importance of taking it into account in teaching. We have also seen that the proportion reporting positive experiences in producing stage 6 questions was larger among those at stage 5 (12 of 22=55%) than those at stage 6 (8 of 9=89%). I have attributed the smaller proportion of stage 6 students to their prior experience in producing stage 6 questions; they had acquired stage 6 Q-forms and it is reasonable to assume that they did not think that producing stage 6 questions was worth reporting.

Q10. If so, what kinds of changes do they notice?

I have shown that most reports involved two of the four target rules, *Cancel-inversion* and *Q-tag*. As I have discussed, this may be because of lack of opportunities to produce *Negative-Aux/Do-2nd* and *Y/N-Negative* questions outside the classroom in daily life in Japan and the fewer class hours allocated to practise these two forms.

I have shown that the reported positive experiences involving *Cancel-inversion* and *Q-tag* ranged over several aspects, such as noticing the use of target rule in someone else's speech in films and conversations, improved performance in using the target rules on the posttests, increased knowledge about the target rules, and experience of starting using the target rules outside the classroom. We have also seen that restructuring of interlanguage can take a long time, and I have pointed out the importance of providing learners with opportunities to use the target rules not only in instruction but in achievement tests, so that learners can realise the usefulness of asking questions by using the target rules.

To sum up, these results show that the noticing-promotion instruction was beneficial in motivating the students ready to learn the target rules to carry on using the target rules beyond the treatment sessions. One can say that the instruction was effective in helping the students become better 'noticers' and users of the target rule in and out of the classroom.

CHAPTER 9
Conclusion

In this final chapter, I will begin by listing the three aims discussed in the previous three chapters. I will then summarise the findings for the 10 research questions in turn. After that, I will discuss limitations of this study. I will then reflect on the whole process of implementing the noticing-promotion approach in the given context, commenting on a number of implications for teaching and for future research.

9.1 Aims of the study

As explained in Chapter 1, my motivation for carrying out this study was my perception of the needs of learners obliged to study English in the classroom, where immediate, intensive teacher feedback on individuals' mistakes is difficult to provide, and also the need for classroom-based research examining the effects of proactive form-focused instruction in communicatively oriented classrooms on individuals' L2 development. Within that larger goal, I attempted to explore two aspects of L2 learning:

(a) the potential and limitations of teachers' planning to provide opportunities for learners to notice the gap between the target grammar and their output through dialogical interaction by providing carefully designed tasks seeded with both the target forms and prompts for learners to assist each other and

(b) learners' perceptions of their gains and usefulness of this type of instruction, which encouraged them to scaffold each other.

To this end, I presented the following three aims:

(a) to examine the effect of the noticing-promotion approach on individuals' sustained gain in developmental stage in English interrogative acquisition;

(b) to examine its effect on individuals' sustained gain in filling gaps in their acquisition of Q-forms targeted in the instruction in relation to their readiness to learn them; and

(c) to explore the students' perceptions of the usefulness of the treatment and of their changes in producing target questions in and out of the classroom.

In the next section, I will summarise the conclusions I have drawn regarding the10 research questions formulated for these three aims.

9.2 Summary of the findings

The findings clearly show that the noticing-promotion approach, in the form of proactive form-focused instruction providing explicit teaching of grammar and noticing tasks, was effective in helping students who were ready to learn the target rules not only to advance in developmental stage in interrogative acquisition but also to fill gaps in their L2 development.

Q1. Is the noticing-promotion approach effective in helping the students advance in developmental stage in interrogative acquisition?

Q2. If so, are there any differences in individuals' sustained stage gain in relation to their current developmental stage?

As explained in Chapter 6, in order to allow a stronger claim that stage advancement had indeed occurred, I used a stricter criterion for developmental stage increase assessment (i.e. at least *two productive* usages of *two different* higher level Q-forms for a given stage on the *two* posttests) than those employed for the developmental stage assessment for the pretest (i.e. at least *two* productive usages of Q-forms for a given stage at *one* time). In Chapter 6, the analysis of student output in the three tests has shown that students who benefited most from the treatment were those at stage 5 on the pretest, who were ready to learn the stage 6 Q-forms; five of the nine stage 5 students (56%) gained a stage on the posttest and maintained it for seven weeks. Moreover, another two stage 5 students progressed in their developmental stage on the follow-up test, demonstrating the delayed effect of the treatment. In contrast, in the comparison group no such positive change (i.e., sustained or delayed effects) was found in any of the three Ready students, and I concluded that the treatment was effective in helping many stage 5 students advance in developmental stage in interrogative acquisition.

I have compared the proportion showing the sustained gain (5 of 9 students=56%) with that observed in

Mackey and Philp's 1998 study (7 of 9 =78%), in which the learners were given immediate, intensive feedback by a trained NS partner in the form of recasts in information-gap tasks without explicit teaching of the target rules, and the same assessment criteria were used. Although in my study the proportion was slightly lower, the period of time over which the students were required to sustain gains in this study was 6-7 weeks – two or three times as long as that in Mackey and Philp (1998) and it is possible to say that the two different approaches had comparable effects. I noted that the sample size in both studies was rather small and we need to be cautious in drawing firm conclusions about the research question, but it is possible to say that the noticing-promotion approach, designed to facilitate learner *noticing* through explicit grammar teaching and noticing tasks in a context where immediate teacher feedback was lacking, is as effective as the recast-rich treatment, provided the learners' next stage of development was targeted in instruction.

Added to this, I have drawn the reader's attention to the fact that in Mackey and Philp (1998) none of the students receiving the reactive FonF instruction advanced to the highest stage. Mackey (1999) explained that this might be due to the fact that the production of stage 6 questions in the instruction was much less frequent than of those at lower stages, and I have pointed out that this is the limitation of this type of instruction. On the other hand, in my study five students advanced to stage 6 after receiving the proactive form-focused instruction, which represents a compromise between FonF and traditional presentation-practice-production instruction, with a strong leaning towards the latter, though not as far as FonF*S* as defined by Long. Many advocates of task-based instruction reject proactive form-focused instruction on doctrinaire grounds, but these results have shown that the noticing-promotion approach used in this study is likely to offer a good chance of success for the systematic teaching of new L2 forms in contexts where time and provision of intensive teacher feedback is limited.

As for the 16 Unready students at stage 4 or below on the pretest, all in the treatment group, only one student showed sustained stage increase (from stage 1 to 3) and one student advanced by two stages (4 to 6) on the posttest and lost one stage and was assigned to stage 5 on the follow-up test. I concluded from these results that the treatment was more beneficial for the students ready to learn the stage 6 rules targeted in the instruction. Although the treatment tasks and tests provided the students with many opportunities to produce not only the target questions but also non-target ones, such as stage 4 and stage

5 questions, these results suggest that the provision of such opportunities alone was not sufficient in helping many students at stage 4 or lower to gain a stage; and that instruction targeted at the next stage of learners' development is necessary.

We have also seen that the treatment was beneficial for the students who had been assigned to the highest stage (i.e. stage 6) on the pretest. Although the stage assessment criterion used for the post- and follow-up tests was more rigorous than that for the pretest, more than half the stage 6 students (12 of 22 students=55%) who received the treatment produced at least two productive examples of two different stage 6 Q-forms on both posttests and maintained stage 6 for seven weeks. By contrast, in the comparison group, only one of the nine students (11%) maintained stage 6 on both posttests. I have concluded that the treatment had positive effects on the production of more stage 6 questions for the stage 6 students.

Q3. Do students show developmental gaps in interrogative acquisition?

As discussed in Chapter 3, the fact that a learner has mastered a processing operation for one stage does not necessarily mean that she/he can at once apply the operations involved in that stage to all the grammatical structures that could be produced, and it was expected that students in this study would show developmental gaps in the acquisition of question formation. In Chapter 7, we have seen that many students actually showed gaps in the acquisition of target forms (*Cancel-Inversion*, *Q-tag*, *Negative-Aux/Do-2nd*) and/or non-target rules (*Y/N-Inversion* and *Wh-Inversion*, both at stage 4) on the pretest, except stage 5 *Aux/Do-2nd*, although no evidence contradicting the 6-stage sequence was observed (i.e. no student skipped a stage).

The gaps were most frequently observed in the acquisition of stage 4 *Wh-inversion* (33 of 47 students =70%), which was not targeted in the treatment, followed by the three target rules, stage 6 *Cancel-inversion* (18 of 31=58%), stage 5 *Negative-Aux/Do-2nd* (21 of 43=49%), and stage 6 *Q-tag* (13 of 31=42%), and by non-targeted stage 4 *Y/N-Inversion* (4 of 47=9%). No gap was observed in the acquisition of the other non-target rules (i.e. stage 5 *Aux/Do-2nd* and those at stage 3 and below).

The gaps for stage 4 *Wh-inversion* were the most frequent in students at different developmental stages including those at the highest stage, and many mistakes in producing *Wh*-inversion questions involved the missing copula (e.g. 'Where the cat?', 'How much the hat?'). As discussed earlier, copula 'be' is considered a variational feature (Pienemann & Johnston, 1986) and I argued that learners' tendency to use a simplified form under communicative demand in information-gap tasks led to this phenomenon. Frequent use of simplified *Wh-inversion* questions can lead learners to fossilization of this rule and I drew the teacher's attention to this problem when providing learners with communicative tasks.

As for the two stage 6 target rules, of the 31 stage 6 students on the pretest, more than half (18 students=58%) showed a gap for *Cancel-inversion* and 13 students (42%) for *Q-tag*. As explained in Chapter 5, many examples of the stage 6 target rules were elicited by the examiner's prompt and it is not possible to draw conclusions about which stage 6 Q-form students find more difficult to learn.

For the two stage 5 rules, whereas half the students (21 of 43=49%) who had reached stage 5 or stage 6 showed a gap for *Negative-Aux/Do-2nd*, none did so for its affirmative form. They produced more examples of *Affirmative-Aux/Do-2nd* in the tests than of *Negative-Aux/Do-2nd*, as is often the case in L2 classes and in real world communication, and I argued that infrequent input and output of *Negative-Aux/Do-2nd* inside and outside of L2 classes may be a reason for the students' frequent failure in applying *Negative-Aux/Do-2nd*. Again, most examples of *Negative-Aux/Do-2nd* were elicited by the prompts seeded in the tests, while many *Affirmative-Aux/Do-2nd* questions were produced voluntarily, and it is not possible to establish, using these results, whether or not there is a difference either in their difficulty of acquisition or their acquisition order. However, the remarkable difference in the number of students showing a gap for these two stage 5 rules suggests a need to provide learners with opportunities to hear and produce *Negative-Aux/Do-2nd* questions in L2 classes to help them acquire this rule. These two Q-forms have been conflated as one Q-form in previous research into the acquisition of question formation, this study is the first to report such a difference in the acquisition of the two rules.

Q4. Is the instruction effective in filling such developmental gaps?

For students to be considered to have acquired a Q-form, they had to produce two productive examples for each Q-form on both posttests. In Chapter 7, we have seen that the treatment was effective in helping more than half students to fill their target rule gaps. For example, in the treatment group, the proportion whose gaps were filled through the treatment was 82 percent (9 of 11 students) for *Cancel-inversion*, 78 percent (7 of 9) for *Q-tag*, and 45 percent (5 of 11) for *Negative-Aux/Do-2nd*. On the other hand, no such striking effect was found in the comparison group; the proportion whose gaps were filled on the posttests was 25 percent (1 of 4) for *Q-tag*, 14 percent (1 of 7) for *Cancel-inversion*, and 13 percent (1 of 8) for *Negative-Aux/Do-2nd*, respectively. These results suggested that the treatment had long-term effects over seven weeks in helping the students fill their developmental gaps.

For the effect of the treatment on filling gaps in the acquisition of non-target Q-forms (i.e. the two stage 4 rules), the treatment was not effective and no such difference was found between the two conditions. In the TG, the proportion of students whose gap was filled through the treatment was 25 percent (1 of 4 students) for *Y/N-inversion* and 16 percent (3 of 19) for *Wh-inversion*. In the CG, gaps were found only for *Wh-inversion* and the proportion of students whose gap for this rule had been filled was 33 percent (2 of 6). As the larger proportion of students' gaps were filled in the acquisition of the target rules, these results suggest that providing opportunities to produce stage 4 questions in the information-gap tasks alone was insufficient to help the students fill these specific gaps even though they had reached stage 4 or higher. This also suggests a need to draw learners' attention to their use of these stage Q-forms, especially to missing copula in *Wh-inversion* questions, through instruction.

While the above four research questions involved analysis of product data from the tests, the following six research questions required analysis of questionnaire data and/or classroom conversation data, used to examine the students' perceptions of usefulness of the noticing-promotion approach, which required them to pay attention to output and to take an active role in assisting their partner to notice and self-correct their mistakes while carrying out the conversation tasks.

Q5. Do the students think the modelling is useful in understanding how to carry out the subsequent dictogloss?

As discussed in Chapter 8, I analysed the students' rating of the usefulness of the modelling video shown in Week 1 on a 5-point scale, from *1* (not useful at all) to *5* (very useful), in relation to their readiness to learn *Q-tag* and the type of modelling video, and I have concluded that both types of video served their purpose as input to show how to carry out the subsequent reconstruction task.. For the two Ready groups, a high rating (i.e. *5* or *4*) was given by more than 70 percent of students ready to learn *Q-tag* (14 of 20 students=70% in the Ready-Peer; 8 of 11=73% in Ready Self), irrespective of the type of modelling video they had watched. Moreover, only three of the 31 Ready students gave a low rating (i.e. *1* or *2*) and the rest of the students in both conditions gave a rating of *3*.

For the two Unready groups, the proportion giving such high rating was larger in the Unready-Peer (6 of 9 =67%) than in the Unready-Self (3 of 7=40%), while a low rating of 1 was given by only one student in the Unready-Peer. The rest of the students (Peer=2, Self=4) gave a rating of *3*. These results suggest that the peer-assistance version was more effective in helping the Unready students better understand how to do the subsequent paired task. I have pointed out the possibility that active interaction shown in the peer-assistance version, such as signalling each other's errors in producing questions and other features and giving hints and/or grammatical explanations, made the content of the conversation salient to many Ready and Unready students and increased understanding of the task procedures. However, the sample size in each condition was small and I refrained from drawing a firm conclusion. More research is needed to examine whether one type of modelling is more effective.

Q6. Do the students think the modelling video is useful regarding any other aspects of L2 learning?

The same questionnaire asked the students to report any benefit of watching modelling in other aspects of L2 learning. Although the main purpose of the modelling video was to help the students understand how to do the subsequent dictogloss reconstruction task in pairs, it was also intended to provide input involving 'the target rule' and 'the non-target L2 items'. Bearing in mind the limitations of using self-reports as evidence, I examined the proportion of students who made a report and the content of the

reported benefit, in relation to the type of modelling video they had watched and their readiness to learn the target rule. In the four conditions, irrespective of the type of modelling video they had watched, more than half of the students commented on the benefit of the modelling video: 13 of the 20 students (65%) in the Ready-Peer, seven of the nine (78%) in the Unready-Peer, nine of the 11 (82%) in the Ready-Self, and four of the seven (57%) in the Unready-Self.

Benefit of the peer-assistance version

For the two groups who watched the peer-assistance version, a similar proportion of students made a report (Ready-Peer=65%, Unready-Peer=78%) and the reported benefits in each group involved three aspects of language use: (a) effective ways of assisting the partner', (b) 'the target rule', and (c) 'the non-target L2 items'. For instance, the reported benefits concerning 'effective ways of assisting the partner' are illustrated by comments in the questionnaire, such as the following:

> (I learnt) how I could help my partner. For example, giving the partner time to reflect on her utterance to make self-corrections is a good way, instead of telling the partner that 'you've made a mistake'.
> by Waka at stage 6

> (I learnt) how to tell the partner that the cause of mistakes was intonation in tag questions.
> by Taka at stage 5

However, for the proportion reporting the benefit for each aspect, there was a difference between the two conditions. For example, the proportion commenting on benefit involving 'effective ways of assisting the partner' was larger in the Unready-Peer (6 of 9 students=67%) than in the Ready-Peer (5 of 20=25%) and I attributed the smaller proportion made by the Ready-Peer group to their familiarity with the seeded input showing good examples of peer assistance, such as 'signalling a mistake to the partner'. They were at a higher developmental stage (i.e. stage 5 or 6) and it is conceivable that the other seeded input, such as 'useful expressions for the management of conversation', was also familiar to many Ready students and that they felt it was not worth reporting them as a benefit.

As for the reported benefits concerning 'the non-target L2 items', the picture was similar; the proportion was also larger in the Unready-Peer (5 of 9=56%) than in the Ready-Peer group (5 of 20=25%) and

again I suggested that the smaller proportion of reports made by the Ready-Peer group was due to their familiarity with the seeded input in the modelling.. The reported benefits involved various aspects, such as (a) 'useful expressions for the management of conversation', (b) 'how to use fillers', and (c) 'pronunciation'.

As for the benefit concerning 'the target rule (*Q-tag*)', the picture was the opposite; whereas only one of the nine students (11%) in the Ready-Peer group made a report, six of the 20 students (30%) did so. The proportion reporting this benefit in the Ready-Peer was smaller than I had expected, but it could be that the students were instructed to learn how to do the reconstruction task through watching the modelling video, rather than to learn the target rule use. In fact, many students who had watched the peer-assistance version (Ready=70%, Unready=67%) gave a high rating of perceived usefulness of this modelling video in understanding how to do the dictogloss reconstruction. The reported benefits included various aspects of the usage of *Q-tag*: (a) 'intonation of tag questions, (b) 'technical terms to talk about the target grammar', and (c) 'noticing possible mistakes in producing tag questions'(2 students).

As regards the lack of report in the Unready-Peer group, I have suggested that it was due to their unreadiness to learn the target rule. These results suggest that what learners can learn from seeded input in modelling is what they are ready to learn. For many Unready students, what was beneficial during watching the peer-assistance version of modelling video was the seeded input involving. 'effective ways of assisting the partner' and 'non-target L2 items' rather than that concerning 'the target rule'. I highlighted the importance of teachers' decisions about the content and amount of input to seed in modelling.

Benefit of the self-correction version

For the self-correction version, the proportion reporting a benefit was larger in the Ready-Self group (9 of 11 =82%) than in the Unready-Self group (4 of 7=57%). The most frequently reported benefit in the two groups watching the self-correction version involved 'the non-target L2 items', such as 'the use of useful expressions for the management of conversation', 'pronunciation', and 'the use of fillers'. This

version did not include scenes of peer assistance, although the videotaped dyad cooperated to complete the reconstruction task, and, unsurprisingly, no student in either condition reported a benefit involving learning 'effective ways of assisting the partner'. Given that 11 of the 29 students (5 Ready and 6 Unready) who had watched the other version commented on this, use of the modelling video including scenes of peer assistance appears to be a useful means of demonstrating taking an active role in peer interaction.

As for a benefit involving 'the use of target rule (*Q-tag*)', few reports were made in either group. For instance, despite the readiness to learn the target rules, none of the 11 Ready students reported a benefit. Since a larger proportion (6 of 20 students) in the other Ready group watching the peer-assistance version modelling made a report, one can say that the peer-assistance version was more effective than the self-correction version in providing comprehensible input concerning the use of 'the target rule' to students ready to learn the rule. As mentioned earlier, the lack of report on the benefit of 'the target rule' was also observed in the other Unready group who had watched the other version of modelling video, suggesting that their unreadiness to learn the target Q-form prevented them from learning from the seeded input.

I drew no firm conclusion from these results due to the small sample size, but suggested the need to include scenes of active peer assistance in modelling in order to make the seeded input more salient to learners. I also suggested the potential of using a modelling video performed by a learner dyad for providing learners with not only comprehensible L2 input but also opportunities to learn effective ways of scaffolding each other in the subsequent task. In L2 classes, modelling is usually given by the teacher and learners have few opportunities to see good models of peer assistance. I also pointed out that, although teachers tend to expect that paired learners will cooperate while doing a task, these results suggested that the benefit of modelling showing smooth conversations, even containing good examples of self-corrections, is rather limited.

Q7. Are there any differences in their *noticing* (i.e. reported successful self-corrections) of Ready students in the subsequent tasks depending on the type of modelling video they have watched?

To compare the effect of the two types of video on the two groups of Ready students' behaviour in the subsequent tasks, I analysed the proportion who reported successful self-corrections made *with* or *without* partner assistance.

Regarding the reported positive experiences in making self-corrections made *without* peer assistance, little difference was found between the two Ready groups in the proportion commenting on 'the target Q-form': eight of 20 students (40%) who had watched the peer-assistant version and four of 11 students (36%) who had watched the other video. For the reported successful self-corrections involving 'the non-target L2 items', the proportion was larger in the Ready-Peer group (3 of 20=15%) than in the Ready-Self group (3 of 11=27%), but the difference was not significant.

For the content of reported self-corrections between the two Ready groups, no substantial difference was observed: comments on 'the target rule' in each group involved three aspects of mistakes in producing or answering tag questions – (a) 'agreement', (b) 'intonation to express certainty of the statement', and (c) 'how to answer'.

As for the successful self-corrections made *with* peer assistance, a similar trend was observed in the proportion of students who made a report concerning 'non-target L2 items': five of the 20 students (25%) in the Ready-Peer group and two of the 11 students (18%) in the Ready-Self group reported positive experiences. On the other hand, for the reports on their positive experiences in making self-corrections involving 'the target rule', the type of modelling video did exert an influence and the proportion of students who made a report was larger in the Ready-Peer group (12 of 20=60%) than in the Ready-Self group (3 of 11=27%). I have suggested the possibility that the provision of opportunities for watching good examples of peer assistance encouraged many Ready students to pay more attention to their partner's output, to scaffold each other to help the partner notice a gap as in the modelling video, and to help the partner self-correct.

For the content of the reported self-corrections made *with* peer assistance, those by 12 of the 20 students (60%) in the Ready-Peer group involved four kinds of mistakes in producing or answering tag questions: (a) 'agreement' which was most often reported (9 of 20=45%), (b) 'intonation' (3=15%), (c) 'how to answer tag questions' (2=10%), and (d) 'misuse of interrogatives in tag questions' (1=5%). On the other hand, those reported by three of the 11 students (27%) in the Ready-Self group involved only two of these four: (b) 'intonation' (2=18%) or (c) 'how to answer' (1=9%), and none of the 11 students reported successful self-corrections involving (a) 'agreement' and (d) 'misuse of interrogatives in tag questions', although many errors in producing tag questions in the treatment and tests involved precisely these aspects. Given that both aspects were focused on in the mini-lecture in the treatment, the lack of positive report on these problematic aspects from Ready students who had watched self-correction suggests it may not be an effective means of promoting mutual scaffolding or leading to gap noticing.

Thus, a larger proportion of students in the Ready-Peer group reported successful self-corrections made *with* assistance, but as many Ready students' gap for *Q-tag* was filled through the treatment (6 of 8=75%) as discussed in Chapter 7, the number who commented on the targeted aspects of use of *Q-tag* was somewhat smaller than expected. I have suggested three reasons. One is that they managed to produce tag questions without mistakes and/or were able to make self-corrections *without* assistance. Another possible reason is that they were not sure whether the self-corrections they had made were correct and hesitated to report them. A third reason is that they were unaware of their successful self-corrections. All this suggests a need to analyse classroom conversation data along with self-reports.

Given my caveat about using students' self-reports on successful self-corrections as evidence that the reported self-corrections *with* assistance *did* occur in the treatment sessions, I presented two illustrative classroom episodes involving self-correcting by two pairs of students (Hide and Mat, and Junko and Fuku). In each episode, one student's mistake in a tag question agreement was noticed by the partner, and the assistance offered to help the partner notice the mistake was effective because it was developmentally sensitive. Patience and repetition of the partner's erroneous utterance, giving his partner time to notice the mistake, eventually helped the partner notice and self-correct.

I also illustrated the limitations of the students' ability to notice errors in producing tag questions in the Memory Game, which was cognitively more demanding than the dictogloss. Fuku, who had acquired *Q-tag*, admitted in a private chat with her partner afterwards that, although she knew she was supposed to pay attention to whether her partner's use of tag questions was appropriate and to help her notice and self-correct mistakes, she could not because she had been preoccupied with needing to reply to her partner's tag questions from memory.

I have concluded from these findings that the peer-assistance modelling video was more effective in facilitating Ready students' noticing and self-correction of the target rule ready for learning in the subsequent tasks, but its effect was influenced by task demands. Further examination of classroom data is needed to draw firm conclusions.

Q8. Are there any differences in their *noticing* (i.e. reported successful self-corrections) of Unready students in the subsequent tasks depending on the type of modelling video they have watched?

For successful self-corrections made *without* peer assistance reported by the two Unready groups, no notable differences were observed in the proportion reporting self-corrections involving 'the non-target L2 items'. A report was made by only one of the nine students (11%) in the Unready-Peer group and two of the seven (29%) in the Unready-Self group.

As for reported successful self-corrections involving 'the target rule (*Q-tag*)', slightly more students in the Unready-Peer group (4 of 9=44%) made a report than those in the Unready-Self group (2 of 7=29%), but the difference was not remarkable. In both conditions, the reported successful self-corrections involved either 'agreement' (3 in the Unready-Peer group and 2 in the Unready-Self group) or 'intonation to express certainty of the statement' in tag questions (2 in the Unready-Peer group and 1 in the Unready-Self group), both aspects of which had been focused on in the mini-lesson given in the treatment.

As for the reported successful self-corrections made *with* peer assistance in producing the target rule (*Q-tag*) in the two Unready groups, only one student in each condition made a report. These students

were not ready to learn the target Q-form and it is conceivable that they were unable to notice the partner's errors in using this rule. It is also possible that Partner A might have been able to signal Partner B's mistakes in target questions, but Partner B was not able to self-correct accordingly. Given that a report was made by the larger proportion of students (12 of 20=60%) in the Ready-Peer group, it is possible to say that learners' readiness had a marked influence on L2 learning.

With regard to reports made on successful self-corrections involving 'the non-target L2 items' (e.g. 'meaning of words', 'misuse of verbs'), the proportion was larger in the Unready-Self group (4 of 9=57%) than in the Unready-Peer group (1 of 7 =11%). I attributed the smaller proportion in the Unready-Peer group to the richer content seeded in the modelling video; in addition to the seeded input involving 'the target rule' and 'the non-target L2 items, they had watched many scenes of active peer assistance in producing tag questions and it is conceivable that this might have drawn their attention excessively to the partner's use of target questions in the subsequent tasks. The reports in these two groups involved either.

In short, these results suggest that what learners can learn from modelling is restricted to what they can process, i.e. what they are ready to learn. If the target form seeded in modelling is beyond the next stage of their development, even if presented in a salient manner such as in the form of peer assistance, such modelling may draw learners' attention to the target form, but can overload them. Learners' attentional resources are limited, and this overloading may prevent them from noticing gaps in their output, both their own and their partner's, and from producing not only what they are ready to learn but what they have already acquired. I have underscored the importance of including the appropriate amount and content of input in modelling.

Q9. Are there any positive changes in the students' use of the target rules beyond the treatment sessions?

In the questionnaire given immediately after the follow-up test, completed 6-7 weeks after the last treatment session, the students were asked to report whether they had detected any positive changes in noticing the use of four target rules (*Cancel-inversion*, *Q-tag*, *Negative-Aux/Do-2nd*, *Y/N-negative*) in

their own or others' speech and in producing one or more target rules beyond the four treatment sessions. Of the 48 students who had received the treatment, nearly half (22=46%) reported positive changes. Such reports were most frequently made by those at stage 5 on the pretest (8 of 9=89%), followed by those at stage 6 (12 of 22=55%), who were ready or had acquired stage 6 Q-forms. Among those who were not ready to learn the stage 6 target rules (N=16), only one stage 3 student made a report. The treatment focused mainly on stage 6 rules and we have seen that it was most beneficial for students at stage 5.

The most frequently reported positive changes in the students' use of the target rules beyond the treatment sessions involved Q-tag and Cancel-inversion, both stage 6 rules; of 27 reports made by the 22 students, 13 concerned *Cancel-inversion* and another 13 involved *Q-tag*. In fact, as shown in Chapter 7, the treatment was effective in helping many Ready students fill gaps for *Q-tag* (6 of 8 students=75%) and *Cancel-inversion* (9 of 11=82%), and these results give support to the possibility that the reported positive changes in using these two stage 6 target rules actually occurred beyond the treatment sessions.

As for stage 5 *Negative-Aux/Do-2nd*, only one report was made on positive changes in its production beyond the treatment sessions, although 50 percent of the 12 students' gaps were filled through the treatment. For stage 6 *Y/N-negative*, no report was made and no data was available regarding the effect of the treatment on filling the gap because its occurrence in the test data was very rare. One possible reason for the lack of comments on these two negative forms of interrogatives is the lack of opportunity to use these rules in daily life in Japan. Another reason could be that the students had a smaller amount of practice in using them in the treatment than the other two Q-forms.

Q10. What kind of positive changes do the students experience beyond the treatment sessions?

Further examination of these reported positive changes beyond the treatment sessions has shown that some students began to notice the use of target rules:

(a) in someone else's speech, such as in talks by an English lecturer in other English courses and in

films; and

(b) in their own production of target questions on various occasions: in the posttests, in other English courses given by a native speaker, and/or at a part-time job, such as teaching English at a cram school and taking an order from a foreigner at a fast-food restaurant.

In addition, some students reported increased knowledge of the target rules through the treatment intended to facilitate noticing gaps, such as the usefulness of *Q-tag* in conversations, and their increased confidence in producing questions with a target Q-form. This is illustrated in the following reports:

> I became more aware of the usage of indirect questions, especially its word order. When I asked questions to my teacher in another course (who was native speaker of English) and asked her to do something for me, I tried to use polite, indirect questions.
> by Nae [Sustained stage gain from 5 to 6]
>
> (After the instruction,) I came to know that the use of tag questions was one of the options for asking questions.
> by Mie [Stage unknown]
>
> I didn't know intonation in the end of tag questions made a difference in meaning until I took the classes. Since then, I began to pay attention to intonation (when producing tag questions). I'm glad I've found out about it.
> by Seigo [Maintained stage 6]
>
> I realised that I was not able to use what I thought I could use, especially those I usually didn't use. It's important to have opportunities to use them through speaking tasks.
> by Saya [Maintained stage 6]

These and other reports show the students' pleasure in their new experiences. EFL learners tend to have few opportunities to use English outside L2 classes, especially late-acquired, cognitively more difficult forms like *Q-tag*, but these reports suggest that the opportunity to learn the target Q-forms through the treatment facilitated gap-noticing and encouraged them to carry on experimenting with the newly learnt form(s) beyond the treatment sessions. I have concluded from these results that the noticing-promotion approach was beneficial in helping many Ready students become better 'noticers' and users of the target rules both inside and outside the classroom. Helping learners to use newly acquired Q-forms for real communication outside class is the ultimate goal of L2 teaching and the positive reports made by these students are encouraging for teachers interested in implementing this type of L2 instruction.

Summary

The findings for the ten research questions have shown the effectiveness of the noticing-promotion approach used in this study in helping a number of Ready learners not only to advance in developmental stage but also to fill gaps in interrogative acquisition. The instruction was also beneficial in encouraging many Ready learners to carry on using the target Q-forms beyond the treatment sessions. These results suggest the significance of this type of instruction in large classes, where immediate teacher feedback on individuals' mistakes is difficult to provide. Above all, the sustained effects observed in the increased developmental stage in this study should be encouraging for teachers, since the treatment effects were similar to those reported in Mackey and Philp's study (1998), where the learners were provided with immediate, intensive recasts from a NS. These results have provided strong support for the use of an approach which consists of explicit teaching of grammar – an element of FonF*S* approaches – followed by proactive FonF activities intended to lead learners from controlled to automatic use of the target structure.

9.3 Limitations of the study

It is important, in considering the results of this study, to note some of its limitations. As discussed in Chapters 5 and 6, in order to allow a stronger claim that stage development had indeed occurred, I used a conservative criterion for analysing individuals' gain in developmental stage (i.e. *two* productive usages of *two* different higher level Q-forms on the *two* posttests), which is the most rigorous among those employed in previous studies to examine the effect of instruction on developmental stage increase in English interrogative acquisition by using the 6-stage sequence. However, the sample size in this study is small and I have been cautious in drawing conclusions.

Likewise, as discussed in Chapters 5 and 7, although the criterion used in the study for analysing the effect of the treatment on filling gaps in the acquisition of Q-forms (i.e. *two* productive usages for each Q-form on the *two* posttests) was more rigorous than 'two productive usages for each Q-form at one time measure' (Pienemann & Johnston, 1987; Pienemann, Johnston, & Brindley, 1988), I have again

exercised caution in my conclusions because I employed emergence-based criteria rather than mastery criteria for the analysis of developmental stage and the acquisition of each Q-form, following Pienemann and Johnston (1987) and others (e.g. Mackey & Philp, 1998). As discussed in Chapter 3, the use of an emergence-based criterion entails the difficulty of distinguishing a formula (i.e. an unanalysed chunk) from an example of successful application of a Q-form. To minimise the scope for such misjudgement, I employed a rigorous criterion for each analysis in order to allow a stronger claim that stage advancement and the acquisition of the target rules had indeed occurred, but any study employing this type of criterion faces this problem.

Another possible limitation is the elicitation methods used to collect target question examples in the two conversation tasks on the three tests. As shown in Chapter 5, a number of target questions were elicited by the prompt, although some higher-level students produced target questions without any cues from the examiner even before the treatment, and many more students did so on the post- and follow-up tests. When students fail to apply an instructed rule on such obligatory occasions, this could be interpreted as evidence for their non-acquisition of this rule in the analysis, but it is conceivable that such a prompt interrupts the conversational flow and draws their attention to *form* while producing questions, and the resulting question examples would be more accurate than those produced without such instructions. Indeed, it would be ideal to collect question examples in natural conversation, such as interviews, but as discussed in Chapters 3 and 5, it is extremely difficult to collect stage 6 question examples through task design without using such instruction. In fact, the production of stage 6 questions in previous studies which did not employ such instruction was rare, and little research has been done to examine the acquisition of the highest-level Q-forms.

In addition, I was keen to carry out my study in the classroom, which inevitably entails less control of variables than in experimental studies. For example, some students performed the tasks with a partner at a higher developmental stage and others with a partner at a lower or the same stage. It is likely that, although the students received the same mini-lesson on the target Q-form and carried out the same noticing tasks, what they experienced through pair work may have varied from student to student. It is also conceivable that what they actually did in and with the same tasks might have differed, since individuals' interest in L2 learning varies, and some might have focused more on the target rule than on

other L2 features, and/or more on fluency than on accuracy, despite the instruction to pay attention to the target form. In this regard, the amount and quality of input and output the students experienced in the four treatment sessions was not identical and this might have influenced their oral performance on the three tests.

The use of students' self-reports on positive experience in producing English in the treatment sessions could be seen as another limitation. As discussed in Chapter 8, it is possible that some students were unaware of successful self-corrections and/or of incorrect self-corrections. Although I have shown some evidence that reported self-corrections did occur in classroom pair work, we have also seen that the reported self-correction did not always take place in the communicative tasks, especially in the cognitively more demanding one (i.e. the information-gap task). There is always a need to analyse the classroom data along with the questionnaire data.

In the wider EFL context, we should bear in mind the homogeneity of the students – they shared the same L1 and had learnt English in the EFL context. This has the advantage of allowing me to eliminate possible L1 influence on L2 production, but has the disadvantage of limiting generalization to other learning contexts – particularly since the range of accepted word order of Q-forms differs from language to language and the rate of acquisition of specific L2 features may be influenced by the L1 (Spada & Lightbown, 1999). More research is needed to examine the effect of this type of instruction with students with a different L1.

9.4 Implications of the study

I will now reflect on a number of implications for teaching, drawn from what this study enabled me to learn about teaching grammar in communicatively-oriented L2 classes to help learners advance in L2 development.

9.4.1 Learners' readiness

The findings clearly showed that the students' readiness played a crucial role in the development of their interrogative acquisition, suggesting the need to focus on structures at the next stage of their development to help them advance, as proposed by Pienemann's Teachability Hypothesis. But many L2 classes contain a large number of students at different developmental stages, especially in the EFL context, and the selection of L2 structures to target is no easy task for teachers.

9.4.2 Developmental gaps

Many students in this study, even those at the highest stage, showed gaps in their L2 development in the acquisition of interrogatives. We have seen that such gaps could be filled through the treatment when it focused on the structures in which learners showed a gap, but that gaps for untargeted Q-forms in the treatment tended not to be filled. Above all, a gap for stage 4 *Wh-Inversion* was frequently observed in students at various developmental levels and their mistakes in producing stage 4 questions often involved a missing copula 'be', a variational feature. I drew teachers' attention to this problem and to the need to help learners notice the problem through *focus on form* instruction.

9.4.3 Acquisition order of interrogatives

The present study was not designed to examine the acquisition order of interrogatives, but the results in fact confirmed that the students followed the 6-stage sequence. One point to note is that no student showed a gap for *Affirmative-Aux/Do-2nd*, while many did for the other stage 5 rule, *Negative-Aux/Do-2nd*. In previous studies these two Q-forms have been conflated as one rule and this

study is the first to report such a difference in the acquisition of the two rules. More research is needed to examine if one form is more difficult to acquire. For the stage 6 Q-forms, no such difference was observed.

9.4.4 The potential of the noticing-promotion approach

I will now summarise the advantages of the noticing-promotion approach used in this study. One is that it can be implemented in large L2 classes, where immediate, intensive teacher feedback on individuals' mistakes is difficult to provide. It requires learner dyads to cooperate with each other and also requires demonstration of procedure through the modelling video. Some may argue that this approach is too demanding for learners, but the present study has shown that this can provide learners with rich *affordance* (i.e. learning opportunities), encouraging them to scaffold each other within their ZPD.

Another advantage of this instruction lies in the use of the modelling video, especially the peer-assistance version. This was appreciated by many students, both Ready and Unready, since it helped them not only to understand how to do the subsequent task but also to learn effective ways of assisting the partner. In addition, the active peer assistance seeded in the modelling seemed to make the target and non-target L2 features more salient to students and so to facilitate their uptake.

The provision of a mini-lesson on the target rule at the beginning of each treatment session seemed effective in drawing the students' attention to it and in providing them with knowledge of its usage, including metalinguistic terms necessary to explain it. Perhaps surprisingly, many students seemed comfortable using these terms once they were explained in the mini-lesson and they watched the modelling video in which the terms were actually used by the learner models. Another strength of providing explicit teaching of grammar is that it can be implemented in courses with a limited number of hours, which other types of approaches such as a rule discovery approach do not so easily allow.

The sequence of noticing tasks also seemed effective in helping students to intake what had been taught in the mini-lesson. The dictogloss was cognitively less demanding and it seemed to allow them to

devote attentional resources to form effectively, while the subsequent information-gap task, which was cognitively more demanding, seemed to have provided them with opportunities to focus more on message conveyance than on form and to use the target rule for communication.

Another issue that emerged is that the treatment encouraged many Ready students to carry on using the target rules beyond the treatment sessions as reflected in their self-reports in the questionnaire on positive changes in noticing. An ultimate goal of L2 teaching is to encourage learners to be autonomous and the findings of my study underline the potential of this approach in this regard.

In conclusion, this study has shown the potential of the noticing-promotion approach as a means of providing rich affordance for L2 development, of encouraging learners to take an active role in scaffolding each other in classroom learning, and of carrying on using the target Q-forms beyond the lessons. Learners in large classes would benefit from this type of instruction but, as I have emphasised, as a complement to the teacher's irreplaceable role in offering feedback on learner mistakes.

9.5 Recommendations for practical applications

In order to increase the likelihood that learners will convert input to intake and will consequently restructure their interlanguage, it would be desirable to focus on one rule in more than one class. For example, it may be helpful if one structure is focused on in two consecutive lessons; in the first lesson learners may receive a mini-lesson on the target rule and carry out two dictogloss tasks, as in this study, including watching a modelling video, followed by an information-gap task more demanding than the dictogloss. And then, instead of performing the information-gap task several times with different task sheets, it may be useful to provide them with a follow-up task to reflect on their output in the task. One beneficial reflection activity would be listening to the recording of their conversation in the information-gap task in order to find mistakes not noticed or incorrectly corrected on-line. In the second lesson, learners could be given a similar information-gap task to the one used in the previous lesson along with other tasks which are intended to draw their attention more to meaning than to form, followed by a post-task, as in the previous lesson.

Although I have suggested focusing on one structure in a lesson to maximise learners' intake, under certain circumstances, it may be beneficial to focus on more than one form in one lesson. For example, if most learners in a class have acquired most stage 6 Q-forms, but tend to ask questions with a familiar, cognitively easier Q-form at stage 5 or lower most of the time, teachers should encourage them to try out Q-forms at different stages, including stage 6 rules, to ask the same thing, by seeding occasions to apply them in noticing tasks. This is because learners' attention tends to be drawn to message conveyance in communication and they tend to miss opportunities to practise unfamiliar, cognitively more difficult Q-forms in L2 classes on their own. In fact, most students in this study, even those who were assigned to stage 6 on the pretest, admitted that they had not produced stage 6 Q-forms for communication either inside or outside the classroom until they took the treatment sessions, although they knew the rules. It is important to help learners learn that they have more than one way of asking the same thing. One tactic is to encourage them to ask the same thing with different Q-forms, followed by a discussion to compare differences among several question examples. In this way, teachers can assist learners to enrich their repertoire.

Another recommendation for practical applications concerns the use of a modelling video. Many students who had watched the peer-assistance version of modelling in this study reported its benefits in learning effective ways of providing feedback to help the partner notice and correct mistakes in producing target questions. In communicatively oriented L2 classes, learners are often paired up and spend a lot of time on various tasks, and helping learners gain confidence about how to provide useful feedback to assist the partner in noticing and correct mistakes may be crucial to enhance peer-mediated learning. The significance of a collaborative peer-peer dialogue in the L2 learning process tends not to be appreciated by learners and it is important that they be aware of why they are being asked to collaborate. One thing that should be noted here is the potential for conflict between learners in taking an active role in providing feedback. In the questionnaires completed in this study, no student commented on any conflict in pair work, but we should be aware of problems which may arise from too many interruptions in conversations by dominant learners.

Another issue involving the use of a modelling video is that it may be useful to stop the video where a

good example is shown and to ask learners what they usually do (or do not do) in such a context, giving them a chance to reflect on their classroom behaviours and possible improvement in their pair work. In this way, teachers can emphasise effective ways of helping the partner notice their mistakes and make self-corrections in the modelling more salient to learners.

It may also be useful to provide reactive FonF activities in lessons after introducing new L2 forms through the noticing-promotion approach for further practice to produce newly introduced forms. In this way, proactive FonF instruction and reactive FonF instruction can complement each other and there would be a better chance of helping learners acquire new forms and advance in L2 development.

9.6 Suggestions for future research

Ideally, it would have been desirable to extend this study to examine whether the noticing-promotion approach brought a longer-term effect in students' L2 development, but this was not feasible due to time constraints and the students' availability.

Another possible extension of the study would be to analyse the classroom conversation data, which would allow us to examine what kinds of noticing actually took place in pair work, whether their *noticing* led students to successful self-correction, and whether there is any relationship between the observed sustained effects and increased noticing and/or successful self-correction. This would also allow us to obtain more precise insights into the potential and limitations of learners' *noticing* and resulting self-correction.

Finally, future research on the use of this type of instruction with students from different L1 backgrounds and in different L2 learning contexts would be needed to find out more about its effect on L2 development in the acquisition of English interrogatives and to see whether the sometimes striking benefits demonstrated by the Japanese learners of English in this study are replicated by their counterparts in other EFL settings.

Notes

[1] Various definitions describing 'mistakes' and 'errors' have been proposed in SLA literature; Corder (1967) defines an 'error' as a deviation in learner language resulting from lack of knowledge of the correct rule and 'mistake' as a deviation in learner language that arises from learners' failure to perform their competence. It is not always easy to distinguish an error from a mistake and in this study the terms are used interchangeably.

[2] There is an ongoing educational reform project to introduce English into elementary schools. A questionnaire survey by the Ministry of Education, Culture, Sports, Science and Technology (MEXT) shows that activities using English were conducted in about 93.6 percent of 22,232 state elementary schools in 2005 (MEXT, 2005. http://www.mext.go.jp/b_menu/houdou/18/03/06031408/001.htm).

[3] In 2001, the former Ministry of Education, Science, Sports and Culture (Monbushō), known as MOE, was merged with the former Science and Technology Agency (Kagaku-Gijutsuchō) and the present Ministry of Education, Culture, Sports, Science and Technology (MEXT) was founded.

[4] Nagasawa (2002) points out two reasons for it: "the shame culture" in Japan, which makes Japanese feel horrified about how other people think of them if they make mistakes; the greater Japanese emphasis on written rather than spoken language in school education; and resultant extreme nervousness when speaking.

[5] In Dekeyser (1995), for example, the subjects were instructed to type sentences of artificial language which they had never encountered in response to colour pictures they had never seen.

[6] Swain's (1998) dictogloss differed from Wajnryb's (1990) in that Swain's did not allow the students to create 'semantic approximations to the original text, created out of the learners' own grammatical and linguistic resources' (Wajnryb, 1990:6), which is one of the features of Wajnryb's dictogloss. It is not clearly stated in Swain (1998), but it seems that this change was intended to make use of the target structures essential to the reconstruction task. Together with explicit teaching of the pre-selected two target rules provided before the dictogloss in each lesson, the treatment was intended to promote metatalk about the target structures in reconstructing the text, which is considered effective for noticing the gap. Nevertheless, although the task objective in Swain (1998) was different from Wajnryb's, the same four-step procedure was adopted to promote the students' active involvement in decision-making about the target language through collaborative conversations, and Swain's activity was a dictogloss in the sense in which Wajnryb uses it. In addition, it should be noted that, while the text in Wajnryb (1990) consists of 4-6 sentences which are textually cohesive with connectives between sentences and reference devices to interconnect ideas, the text used for the treatment in Swain (1998) was shorter (two sentences) except for one employed for data collection. It seems that use of the shorter text was due to the time constraint and the subjects' young age. See also note 23 for use of a different type of text for a dictogloss.

[7] In addition to interrogatives, Pienemann and Johnston (1986, 1987) listed other English features, such as dative movement (stage 6) as in 'I gave John a gift', third person singular '-s' marking (stage 5), and insertion 'to' as a complementiser (stage 4) as in 'want to go'.

[8] In Spada and Lightbown's 1999 study, the number of students who increased in developmental stage on the pre- and posttests was counted, but their 1993 study had examined the number of students who produced questions at each stage on each test. So direct comparison in terms of the number of individuals who moved up a stage is not possible.

[9] Spada and Lightbown (1999) employed both conversation and written tasks in data collection and

found that the learners' performances on the oral and written tasks resulted in different stage assignments, leading to higher stage assignments in the oral task than in the written task.

[10] It should be noted that, in this study, learner' changes in their L2 development was only considered in relation to question formation and no claims about the overall developmental stage of learners' interlanguage were made.

[11] In Japan, a new academic year runs from April to the end of March at most universities and schools. The study was carried out at the two institutions in the second semester of the same academic year.

[12] Two students in Class B were absent when my representative visited the class. A copy of my letter was given by the lecturer to each student later and they also signed up for my study.

[13] Original copies of the completed Consent Form taken from the 59 students included in the analysis are available on request.

[14] For the nursing college students, the *English 1 Course* in which I used the materials developed for the present study was a compulsory course and I did not need to prepare a letter to invite their participation.

[15] At the university, many students were interested in teaching English to school children and I prepared a listening task to teach linking sounds by using a nursery song called 'Pat-a-Cake, Pat-a-Cake' and then taught how to dance along with the words. I used the same material at the nursing college. In addition, at the nursing college I had two extra classes after the follow-up test and used the textbook *Christine's Easy English Conversation for Nursing* (Chinen & Kamitaki, 1995) to teach basic conversations.

[16] I was on suspension twice in my PhD study due to family reasons, which totalled 15 months.

[17] In Fountain & Nation (2000), only the underlined 20 key words in each main paragraph were given points. One mark was given for each key word, making a possible maximum of 80 marks for the whole test.

[18] One week before the two tests were administered in one of the classes for the course the students were enrolled in, the students were encouraged by the course director to join the researcher's study and told that, if they wished to participate, they had to take the two tests given in the next class.

[19] For two students in Class B who were absent this day (Toby and Kada), I gave the Dictation Test and the questionnaire about two weeks before the pretest. They did not take the Grammaticality Judgment Test because it was intended to prepare the materials for teaching and testing and I had finished their preparation.

[20] This information was given to me by the research assistant.

[21] In private communication in the main study, she told me that she felt extremely tired after the Dictation Test in which she had to listen to the recorded text spoken at natural speed. She was reluctant to do the Grammaticality Judgement Test and ticked 'I don't know' without trying to read the given 50-pairs of sentences.

[22] At the end of the first treatment session, I asked the students in each class whether or not they had taken English courses which focused on the use of any question forms as in my class. No student had.

[23] The text used in this study also differed from that used by others (e.g., Wajnryb, 1990; Swain, 1998) in that the latter is continuous monologue, written in the form of a paragraph, while that used in my study consisted of questions and answers with some discourse markers and linkers produced by two native

speakers as they carried out the information gap task. Nevertheless, the same four-step procedure was adopted to promote the students' active involvement in decision-making about the target language through collaborative conversations, and my activity was a dictogloss in Wajnryb's sense in terms of the steps followed, although it differed somewhat in terms of my purpose. As described above, my activities came closer to FonF*S* than to FonF, as it was the case in Swain (1998); and I was concerned with accurate grammatical production rather than textual cohesion. It is interesting to point out that the interactive spoken texts used in my study were seeded with the discourse markers and linkers and many of them were picked up by a number of students in the subsequent tasks, although the students were not asked to include those discourse features in their reconstructed texts.

[24] A copy of the modelling video used in the four treatment sessions is available on request.

[25] 'GUESS WHO?' is manufactured by Hasbro International Inc. I purchased eight packages of this game for the present study and borrowed five from my friends. In each package, two playing boards are provided to place a set of 24 portraits upright facing the player, but I did not use these boards for two reasons. First, if the boards are used, lots of noise is made as the students flip down cards – a distraction. Second, each package is bigger than an A-4 book and is difficult to carry around.

[26] This question could be asked only once in one game by each student in pairs.

[27] *Wh-Inversion* was not one of the targets of the instruction, but I included this example in the handout used in Week 3. As reported earlier, the results of the preliminary tests showed that many students at the nursing college might not have reached stage 4 or above and it was hoped that inclusion of the stage 4 question example would help weaker students.

[28] I should have included an example of mistakes involving the missing phenomenon in the subordinate clause mentioned in the grammar lesson.

[29] It is possible to use the other negative Q-form (*Y/N-Negative*) to press invitations and offers, as in '*Won't you* come in for a few minutes?', but I did not include such examples in the mini-lesson because I intended to draw the students' attention to the differences between the two negative Q-forms.

[30] After the two conversation tasks on each test, the students were given a monologue task, called the Picture-cued Production, on each test. This task was added to collect more data in anticipation of a potential problem which would have been caused by the non-availability of one set of participants at the university. This task was intended to measure the students' ability to generate questions orally without a conversation partner, allowing them to pay attention to forms with less communicative stress, and for me to compare each student's use of Q-forms in the monologue task with that in the conversation tasks. However, the question examples produced in this task were excluded from the analysis of this study, since the problem did not materialise. Another reason for excluding these question examples is that learners' interlanguage is best represented in examples produced in the relatively spontaneous and free oral production tasks, as discussed in Section 5.8.1.

[31] When the Japanese talk with someone older than they are, they tend to feel obliged to use polite forms of words and expressions, even if they have stayed with them before. I hoped that this tendency would help me collect examples of *Cancel-Inversion* in the Guided Role-play.

[32] The examiners were instructed to give a positive comment on the reported number of differences by the students to encourage them to keep on with the task.

[33] One of the two audio recorders was a minidisk and the other was either a cassette or a minidisk.

[34] I got help from one of the examiners to transcribe a part of the recordings. She transcribed one third taken in one of the three tasks (Picture-cued Production) and I double checked her transcripts.

[35] Spada and Lightbown (1993) broke down Stage 3 into various subcategories to reflect behaviours observed in questions produced by young Francophone learners, but their coding system was the same as that described in Pienemann and Johnston (1986). They distinguished in stage 3 between *Do-fronting* and *Fronted-other* categories, where learners showed evidence of the rule application of fronting, but where the fronted element was not 'do' but some other auxiliary verbs, as in 'Is the ball is in the air?'. Within *Fronted-other* category, they further distinguished *wh-fronted* questions *with* the inclusion of an auxiliary verb (e.g. What color the dog?) and *without* the inclusion of an auxiliary verb (e.g. What color the dog is?). Such examples were also observed in the coding process of this study and I partially adopted Spada and Lightbown's *Front other* category. For *wh-fronted* questions *with* the inclusion of an auxiliary verb, I coded them as stage 3 questions, but for *wh-fronted* questions *without* the inclusion of an auxiliary verb, I did not code them as stage 3 questions but as stage 1 questions. This was because *wh-fronted* questions *without* the inclusion of an auxiliary verb were more like *sentence fragments* than productive examples of the rule application of *Fronting*. On the other hand, in *wh-fronted* questions *with* the inclusion of an auxiliary verb, we can identify the rule application of *Canonical word order* (stage 2), which is the prerequisite for the stage 3 rules with a *Fronting* element.

[36] In general, the students seemed curious about the content of each modelling version, especially in the first two treatment sessions, but in Week 3, I observed a few students yawning while watching each version of video. One reason for this is the content of the interaction of the modelling video used in each session was basically similar, except for the targeted Q-forms, and it is possible that predictability led to boredom.

[37] As noted in Chapter 5, some students' test performances were not recorded for technical reasons. Their data was excluded in the analysis of developmental stage, but I included one student's report on her positive experience in producing examples of target rule in this analysis and indicated her developmental stage as 'unknown'. This student was at the nursing school and from my observation in the instruction, it seems that she had not acquired this rule before the instruction and that her developmental stage was stage 4 or below.

[38] In the questionnaire, the students were asked to report positive changes in using the target rules after the instruction, but no questions were included to assess improvement in other areas of L2 use. So the lack of reports by lower-level students does not necessarily mean that they did not notice any positive changes.

[39] Mie's performance on the pretest was not recorded and it was not possible to examine her *Q-tag* acquisition. From my observation in the treatment at the nursing school, it seems that she had not acquired this rule before the instruction.

References

Books and Journal articles

Adam, M. (1978). Methodology for examining second langue acquisition. In Hatch, E. (ed.) *Second Language Acquisition: a book of reading*. Rowley, MA: Newbury House.

Aljaafreh, A & Lantolf, J. (1994). Negative feedback as regulation and second language learning in the Zone of Proximal Development. *The Modern Language Journal*, 78.4: 465-483.

Allen, L. Q. (2000). Form-meaning connections and the French causative: an experiment in processing instruction. *Studies in Second Language Acquisition*, 22.1: 69-84.

Anderson, J. (1987). Skill Acquisition: Compilation of weak-method problem solutions. *Psychological Review*, 94.2: 192-210.

Anderson, J. (1993). *Rules of the Mind*. Hillsdale, NJ: Lawrence Erlbaum.

Anderson, J. (1995). *Learning and Memory: An Integrated Approach*. New York: Wiley.

Bailey, N., Madden, C. & Krashen, S. (1974). Is there a "natural sequence" in adult second language learning. *Language Learning*, 24.2: 235-243.

Bialystok, E. (1979). An analytical view of second language competence: a model and some evidence. *The Modern Language Journal*, 63.5-6: 257-262.

Bialystok, E. (1982). On the relationship between knowing and using forms. *Applied Linguistics*, 3.3: 181-206.

Bialystok, E. (1991). Metalinguistic dimensions of bilingual language proficiency. In Bialystok, E. (ed.) *Language Processing in Bilingual Children*, 113-140. Cambridge: CUP.

Bialystok, E. (1994). Analysis and control in the development of second language proficiency. *Studies in Second Language Acquisition*, 16.2: 157-168.

Brown, P. C. (2001). The interactive dictation. *The Language Teacher* Online, 25.07.
URL: http://www.jalt-publications.org/tlt/articles/2001/07/brown

Brown, R. (1973) *A first language: the early stages*. Cambridge, MA: Harvard University Press.

Bygate, M., Skehan, P., & Swain, M. (eds.) (2001) *Researching pedagogic tasks: second language learning, teaching and testing*. London: Longman.

Cancino, H., Rosansky, E, and Schuman, J. (1978). The acquisition of English negatives and interrogatives by native Spanish speakers. In Hatch, E. (ed.) *Second Language Acquisition: a book of reading*. Rowley, MA: Newbury House.

CAG-no-kai (1993). *80 Communication Games for Japanese Language Teachers*. Tokyo: Japan Times

Carroll, S. & Swain, M. (1993). Explicit and implicit negative feedback: an empirical study of the learning of linguistic generalizations. *Studies in Second Language Acquisition*, 15.3: 357-386.

Chaudron, C. (1985). A method for examining the input/intake distinction. In Gass, S & Madden, C. (eds.) *Input in Second language acquisition*. Rowley, MA: Newbury House.

Clennell, C. (1999). Promoting pragmatic awareness and spoken discourse skills with EAP classes. ELT Journal, 53.2: 83-91.

Coughlan, R. & Duff, P. (1994). Same task, different activities: analysis of a second language acquisition task from an activity theory perspective. In Lantolf, J.P. & Appel, G. (eds.) *Vygotskian approaches to second language research*, 173-194. Norwood, NJ: Ablex Press.

Dekeyser, R. (1995). Learning L2 grammar rules: an experiment with a miniature linguistic system. *Studies in Second Language Acquisition*, 17: 379-410.

Dekeyser, R. (1998). Beyond focus on form: cognitive perspectiveson learning and practicing second language grammar. In Doughty, C. & Williams, J. (eds.) *Focus on Form in Classroom Second Language Acquisition*, 42-63. Cambridge: CUP.

DeKeyser, R. M. & Sokalski, K. J. (1996). The differential role of comprehension and production practice. *Language Learning*, 46.4: 613-642.

Donato, R. (1994). Collective scaffolding in second language learning. In Lantolf, J.P. & Appel, G. (eds.) Vygotskian approaches to second language research, 33-56. Norwood, NJ: Ablex Press.

Doughty, C. (1988). The effects of instruction on the acquisition of relativization in English as a second language. Unpublished doctoral dissertation, University of Pennsylvania, Philadelphia.

Doughty, C. (1991). Second language instruction does make a difference: evidence from an empirical study of SL relativization. *Studies in Second Language Acquisition*, 13.4: 431-470.

Doughty, C. (1994). Fine-tuning of feedback by competent speakers to language learners. In Alatis (ed.) *GURT 1993: Strategic interaction*, 96-108. Washington, DC.: Georgetown University Press.

Doughty, C., & Pica, T. (1986). "Information gap" tasks: Do they facilitate second language acquisition? *TESOL Quarterly*, 20.2: 305-325.

Doughty, C. & Williams, J. (1998). Pedagogical choices in focus on form. In Doughty, C. & Williams, J. (eds.) *Focus on Form in Classroom Second Language Acquisition*, 197-261. Cambridge: CUP.

Dulay, H. & Burt, M. (1973). Should we teach children syntax? *Language Learning*, 23.2:245-258

Eckman, F., Bell, L. and Nelson, D. (1988). On the generalization of relative clause instruction in the acquisition of English as a second language. *Applied Linguistics*, 9.1: 1-20.

Ellis, N. (1993). Rules and instances in foreign language learning: Interactions of explicit and implicit knowledge. *The European Journal of Cognitive Psychology*, 5.3: 289-318.

Ellis, N. (ed.) (1994). *Implicit and explicit learning of languages*. London: Academic Press.

Ellis, R. (1985). *Understanding Second Language Acquisition*. Oxford: OUP.

Ellis, R. (1990). *Instructed Second Language Acquisition*. Oxford: Blackwell.

Ellis, R. (1994). *The Study of Second Language Acquisition*. Oxford: OUP.

Ellis, R. (1997). *Second Language Acquisition*. Oxford: OUP.

Ellis, R. (2003). *Task-Based Language Learning and Teaching*. Oxford: OUP.

Ellis, R., Basturkmen, H. & Loewen, S. (2001). Learner uptake in communicative ESL lessons. *Language Learning*, 51.2: 281-318.

Erlam, R. (2003). Evaluating the relative effectiveness of structured-input and output-based instruction in foreign language learning: Results from an experimental study. *Studies in Second Language Acquisition*, 25.4: 559-582.

Felix, S. (1981). The effect of formal instruction on second language acquisition. *Language Learning*, 31.1: 87-112.

Foster, P. (1998). A classroom perspective on the negotiation of meaning. *Applied Linguistics*, 19.1: 1-23.

Fotos, S. (1993). Consciousness raising and noticing through focus on form: Grammar task performance versus formal instruction. *Applied Linguistics*, 14.4: 385-407.

Fotos, S. (1994). Integrating grammar instruction and communicative language use through grammar consciousness-raising tasks. *TESOL Quarterly*, 28.2: 323-351.

Fountain, R. & Nation, I. (2000). A vocabulary-based graded dictation test. *RELC Journal*, 31.2: 29-44.

Frawley, W. & Lantolf, J. P. (1985). Second language discourse: a Vygotskyan perspective. *Applied Linguistics*, 6.1: 19-44.

Garciá, P. & Asención, Y. (2001). Interlanguage development of Spanish learners: Comprehension, production, and interaction. *The Canadian Modern Language Review/La Revue canadienne des langues vivantes,* 57: 377-401.

Gass, S. (1982). From theory to practice. In Hines and Rutherford (eds.) *On TESOL 1981*, 36-42. Washington, D. C.: TESOL.

Gass, S. (1988). Integrating research areas: a framework for second language studies. *Applied Linguistics*, 9.2:198-217.

Gass, S., Mackey, A. & Pica, T. (1998). The role of input and interaction in second language acquisition introduction to the special issue. *The Modern Language Journal*, 82.3: 299-307.

Gass, S. & Varonis, E. (1985). Task variation and nonnative/nonnative negotiation of meaning. In Gass, S. and Madden, C. G. (eds.) *Input in Second Language Acquisition*, 235-253. Rowley, MA: Newbury House.

Gass, S. & Varonis, E. (1994). Input, interaction, and second language production. *Studies in Second Language Acquisition*, 16.3: 283-302.

Harley, B. (1992). Patterns of second language development in French immersion. *Journal of French Language Studies*, 2.2: 159-183.

Harley, B. & Swain, M. (1984). The interlanguage of immersion students and its implications for second language teaching. In Davies, A. Criper, C. & Howatt, A.P.R. (Eds.) *Interlanguage*. 291-311. Edinburgh: Edinburgh University Press.

Hatch, E. (1978). Apply with caution. *Studies in Second Language Acquisition*, 2.1: 123-143.

Havranek, G. & Cesnik, H. (2001). Factors affecting the success of corrective feedback. *EUROSLA Yearbook*, 1.1: 99-122.

Hino, N.(1988). Yakudoku: Japan's dominant tradition in foreign language learning. *JALT Journal*, 10.2: 45-55.

Holliday, A. (1994). *Appropriate Methodology and Social Context*. Cambridge: Cambridge University Press.

Howatt, A.P.R. (1984). *A History of English Language Teaching*. Cambridge: Cambridge University Press.

Hudson, T. (1993). Nothing does not equal zero: Problems with applying developmental sequence findings to assessment and pedagogy. *Studies in Second Language Acquisition*, 15.4: 461-493.

Hulstijin, J. (1987). Onset and development of grammatical features: two approaches to acquisition orders. Paper presented at Interlanguage Conference, Trobe University, Melbourne.

Jacobs, G. (2003). Combining dictogloss and cooperative learning to promote language learning. *The Reading Matrix*, 3.1: 1-15.

Johnston, M. (1985). *Syntactic and morphological progressions in learner English*. Canberra: Department of Immigration and Ethnic Affairs.

Johnston, M. (1987). Second language acquisition research in the adult migrant education program. In Johnston, M. & Pienemann, M. (eds.) *Second Language Acquisition: a Classroom Perspective*, 100-125. New South Wales Migrant Education Service.

Kellerman, E. (1985) If at first you do succeed… In Gass, S. & Madden, C. (eds.) *Input in Second Language Acquisition*, 345-353. Rowley, MA: Newbury House.

Kim, S. (2001). *Structured-input and production practice in foreign/second language learning*. Unpublished doctoral dissertation. Temple University Japan, Tokyo.

Kowal, M., & Swain, M. (1994). Using collaborative language production tasks to promote students' language awareness. *Language Awareness*, 3.2: 73-93.

Kowal, M. & Swain, M. (1997). From semantic to syntactic processing: how can we promote it in the immersion classroom? In Johnson, R. & Swain, M. (eds.) *Immersion Education: international Perspectives*, 284-309. Cambridge: CUP.

Krashen, S. (1980). The input hypothesis. *Georgetown University Round Table on Languages and Linguistics*, 168-180. Washington, D.C.:Georgetown.

Krashen, S. (1981). The "fundamental pedagogical principle" in second language teaching. *Studia Linguistica*, 1981, 35.1-2: 50-70.

Krashen, S. (1982). *Principles and Practice in Second Language Acquisition*. New York: Pergamon.

Krashen, S. (1985). *The Input Hypothesis*. Torrence, CA: Laredo.

Krashen, S. (1992). "Teaching issues: Formal grammar instruction" Two readers react. *TESOL Quarterly*, 26.2: 409-411.

Krashen, S. (1994). The pleasure hypothesis. *Georgetown University Round Table on Languages and Linguistics*, 299-322.

Krashen, S. & Terrell, T. (1983). *The natural approach: Language acquisition in the classroom.* CA: The Alemany Press: Hayward.

Lantolf, J. P. (1994). Sociocultural theory and second language learning: introduction to the special issue. *The Modern Language Journal,* 78.4: 418-420.

Lantolf, J. P. (1996). SLA theory building: "letting all the flowers bloom!". *Language Learning*, 1996, 46.4: 713-749.

Lantolf, J. P. (2000a). Second language learning as a mediated process. *Language Teaching*, 33: 79-96.

Lantolf, J. P. (2000b). Introducing sociocultural theory. In Lantolf, J.P. *Sociocultural Theory and Second Language Learning,* 1-26. Oxford: OUP.

Lantolf, J. P. & Aljaafreh, A. (1995). Second language learning in the Zone of Proximal Development: a revolutionary experience. *International Journal of Educational Research*, 23.7: 619-632.

Lantolf, J.P. & Appel, G. (1994). Theoretical framework: an introduction to Vygotskian approaches to second language research. In Lantolf, J.P. & Appel, G. (eds.) *Vygotskian approaches to second language research*, 1-32. Norwood, NJ: Ablex Press.

Lapkin, S. & Swain, M. (1996). Vocabulary teaching in a grade 8 French immersion classroom: a descriptive case study. *The Canadian Modern Language Review/La Revue canadienne des langues vivantes,* 53.1: 242-256.

Larsen-Freeman, D. & Long, M. (1991). *An Introduction to Second Language Acquisition Research.* London: Longman.

Leow, R. (1997). Attention, awareness, and foreign language behavior. *Language Learning*, 47.3: 467-505.

Lightbown, P. M. (1998). The importance of timing in focus on form. In Doughty, C & Williams, J. (eds.) *Focus on Form in Classroom Second Language Acquisition*, 177-196. Cambridge: CUP.

Lightbown, P. M. (2000). Classroom SLA research and second language teaching. *Applied Linguistics*, 21.4: 431-462.

Loewen, S. (2005). Incidental focus on form and second language learning. *Studies in Second Language Acquisition*, 27.3: 361-386.

Long, M. (1980). Inside the "black box": Methodological issues in classroom research on language learning. *Language Learning*, 30.1: 1-42.

Long,, M. (1983). Does second language instruction makes a difference? A review of the research. *TESOL Quarterly*, 17.3: 359-382.

Long, M. (1985). Input and second language acquisition theory. In Gass, S. & Madden, C. G. (eds.) *Input in Second Language Acquisition*, 377-393. Rowley, MA: Newbury House.

Long, M. (1991). Focus on form: a design feature in language teaching methodology. In de Bot, K, Ginsberg, R. & Kramsch, C. (eds.) *Foreign Language Research in Cross-cultural Perspective*,

39-52. Amsterdam: John Benjamins.

Long, M. (1996). The role of the linguistic environment in second language acquisition. In Ritchie, W. C. & Bhatia, T. K. (eds.) *Handbook of Language Acquisition. Vol. 2: Second Language Acquisition*, 413-468. New York: Academic Press.

Long, M & Robinson, P. (1998). Focus on form: theory, research and practice. In Doughty, C. & Williams, J. (eds.) *Focus on form in classroom language acquisition*, 15-41. Cambridge: CUP.

Loschky, L. (1994). *Comprehensible* Input and Second Language Acquisition: What Is the Relationship? *Studies in Second Language Acquisition*, 16.3: 303-323.

Loschky, L. & Bley-Vroman, R. (1993). Grammar and task-based methodology. In Crookes, G. & Gass, S. (eds.) *Tasks and Language Learning: Integrating Theory and Practice*, 123-167. Clevedon: Multilingual Matters.

Lynch, T. (2001). Seeing what they meant: transcribing as a route to noticing. *English Language Teaching Journal*, 55.2: 124- 132.

Lynch, T. (2005). Self-transcribing and noticing in EAP speaking classes. *Edinburgh Working Papers in Applied Linguistics*, 14: 54-67.

Lynch, T. & Maclean, J. (2000). Exploring the benefits of task repetition and recycling for classroom language learning. *Language Teaching Research*, 4.3: 221-250.

Lynch, T. & Maclean, J. (2001). 'A case of exercising': Effects of immediate task repetition on learners' performance. In Bygate, M., Skehan, P., Swain, M. (eds.) *Researching Pedagogic Tasks: Second Language Learning, Teaching and Testing*. Harlow, Essex: Longman.

Lyster, R. (1994). The effect of functional-analytic teaching on aspects of French immersion students' sociolinguistic competence. *Applied Linguistics*, 15.3: 263-287.

Lyster, R. (1998a). Immersion pedagogy and implications for language teaching. In Cenox, J & Genesee, F. (eds.) *Beyond bilingualism: Multilingualism and multilingual education*, 64-95. Clevedon: Multilingual Matters.

Lyster, R. (1998b). Negotiation of form, recasts, and explicit correction in relation to error types and learner repair in immersion classrooms. *Language Learning*, 48.2: 183-218.

Lyster, R. (1998c). Recasts, repetition, and ambiguity in L2 classroom discourse. *Studies in Second Language Acquisition*, 20.1: 51-81.

Lyster, R. & Mori, H. (2006). Interactional feedback and instructional counterbalance. *Studies in Second Language Acquisition*, 28.2: 269-300.

Lyster, R. & Ranta, L. (1997). Corrective feedback and learner uptake: Negotiation of form in communicative classrooms. *Studies in Second Language Acquisition*, 19.1: 37-66.

Mackey, A. (1995). *Stepping up the pace: Input, interaction and interlanguage development. an empirical study of questions in ESL*. Unpublished doctoral dissertation, University of Sydney, Sydney.

Mackey, A. (1999). Input, interaction, and second language development: an empirical study of question formation in ESL. *Studies in Second Language Acquisition*, 21.4: 557-587.

Mackey, A., Gass, S. & McDonough, K. (2000). How do learners perceive interactional feedback? *Studies in Second Language Acquisition*, 22.4: 471-497.

Mackey, A. & Oliver, R. (2002). Interactional feedback and children's L2 development. *System*, 30.4: 459-477.

Mackey, A. & Philp, J. (1998). Conversational interaction and second language development: recasts, responses, and red herrings? *The Modern Language Journal*, 82.3: 338-356.

McConell, D. L. (2000). *Importing Diversity: inside Japan's JET Program*. California: University of California Press.

McDonough, K. (2005). Identifying the impact of negative feedback and learners' responses on ESL question development. *Studies in Second Language Acquisition*, 27.1: 79-103.

McLaughlin, B. (1978). The monitor model: some methodological considerations. *Language Learning*, 28.2: 309-332.

McLaughlin, B. (1987). *Theories of Second-language Learning*. London: Edward Arnold.

McLaughlin, B. (1990). Restructuring. *Applied Linguistics*, 11.2: 113-128.

McLaughlin, B. & Heredia, R. (1996). Information-processing approaches to research on second language acquisition and use. In Ritchie, W. & Bhatia, T. (eds.) *Handbook of Second Language Acquisition*, 213-28. London: Academic Press.

Meisel, J., Clahsen, H. & Pienemann, M. (1981). On determining developmental stages I natural second language acquisition. *Studies in Second Language Acquisition*, 3.2: 109-135.

Mennim, P. (2005). Noticing tasks in a university EFL presentation course in Japan: their effects on oral output. Unpublished doctoral thesis. University of Edinburgh, Edinburgh.

MEXT (2005). Kyoiku Shihyo no Kokusai Hikaku/The International Comparison of the Index of School Education.
URL: http://www.mext.go.jp/b_menu/houdou/18/03/06032718/001/002.pdf

MEXT (2006). *The Course of Study for Lower Secondary School.*
URL: www.mext.go.jp/english/ shotou/030301.htm

MEXT (2006). *The Course of Study for Upper Secondary School*
URL: www.mext.go.jp/english/shotou/030301.htm

Morris, F. & Tarone, E. (2003). Impact of classroom dynamics on the effectiveness of recasts in second language acquisition. *Language Learning*, 53.2: 325-368.

Nagasawa, K. (2002). Teacher Training and development. In Makarova. V. & Rogers, T. (eds.) *English Language Teaching: the case of Japan*, 280-295. München : LINCOM Europa.

Newman, F. & Holzman, L. (1993). *Lev Vygotsky: Revolutionary Scientist*. New York, NY: Routledge.

Norris, J. & Ortega, L. (2000). Effectiveness of L2 instruction: a research synthesis and quantitative meta-analysis. *Language Learning*, 50.3: 417-528.

Ohta, A. (2000). Rethinking interaction in SLA: Developmentally appropriate assistance in the Zone of Proximal Development and the acquisition of L2 grammar. In Lantolf, J. P. *Sociocultural Theory*

and Second Language Learning, 51-78. Oxford: OUP.

Oller, J. (1979). *Language tests at school*. London: Longman.

Panova, I. & Lyster, R. (2002). Patterns of corrective feedback and uptake in an adult ESL classroom. *TESOL Quarterly*, 36.4: 573-595.

Philp, J. (2003). Constraints on "noticing the gap": Nonnative speakers' noticing of recasts in NS-NNS interaction. *Studies in Second Language Acquisition*, 25.1: 99-126.

Pica, T. (1987). Second-language acquisition, social interaction, and the classroom. *Applied Linguistics*, 8.1: 3-21.

Pica, T. (1988). Interlanguage adjustments as an outcome of NS-NNS negotiated interaction. *Language Learning*, 38.1: 45-73.

Pica, T. & Doughty, C. (1985). The role of group work in classroom second language acquisition. *Studies in Second Language Acquisition*, 7. 2: 233-486.

Pica, T., Kanagy, R. & Falodun, J. (1993). Choosing and using communication tasks for second language instruction. In Crookes, G & Gass, S. (eds.) *Tasks and Language Learning: Integrating Theory & Practice*, 9-34. Clevedon: Multilingual Matters.

Pica, T., Young, R., & Doughty, C. (1987). The impact of interaction on comprehension. *TESOL Quarterly*, 21.4: 737-758.

Pienemann, M. (1984). Psychological constraints on the teachability of languages. *Studies in Second Language Acquisition*, 6.2: 186-214.

Pienemann, M. (1986). Is language teachable? Psycholinguistic experiments and hypotheses. *Australian Working Papers in Language Development*, 1: 1-41.

Pienemann, M. (1987). Determining the influence of instruction on L2 speech processing. *Australian Review of Applied Linguistics*, 10.1: 83-113.

Pienemann, M, (1988). Determining the Influence of Instruction on L2 Speech Processing. In: G. Kasper (ed.): Classroom Research. *AILA Review*, 5: 40-72

Pienemann, M. (1989). Is language teachable? Psycholinguistic experiments and hypothesis. *Applied Linguistics*, 10.1: 52-79.

Pienemann, M. (1992). COALA-A A computational system for interlanguage analysis. *Second Language Research*, 8: 59-92.

Pienemann, M. (1998). *Language processing and second language development: Processability theory*. Amsterdam: John Benjamins.

Pienemann, M. & Johnston, M. (1986). An acquisition based procedure for second language assessment (ESL). *Australian Review of Applied Linguistics*, 9.1: 92-122.

Pienemann, M. & Johnston, M. (1987). Factors influencing the development of language proficiency. In Nunan, D. (ed.) *Applying Second language Acquisition Research*, 45-141. Adelaide, Australia: National Curriculum Resource Centre, AMEP

Pienemann, M., Johnston, M. & Brindley, G. (1988). Constructing an acquisition-based procedure for second language assessment. *Studies in Second Language Acquisition*, 10: 217-243.

Pienemann, M. & Mackey, A. (1993). An empirical study of children's ESL development. In McKay, P.(ed.) *ESL Development: Language and Literacy in Schools. Vol. 2: Documents on Bandscale Development and Language Acquisition*, 115-259. Canberra: National Languages & Literacy Institute of Australia and Commonwealth of Australia.

Robinson, P. (1996). Learning simple and complex second language rules under implicit, incidental, rule-search, and instructed conditions. *Studies in Second Language Acquisition*, 18.1: 27-67.

Ravem, R. (1974). The development of WH-questions in first and second languae learners. In Schumann, J & Stenson, N. (eds.) *New Frontiers in Second Language Learning*. Rowley, MA.: Newbury House.

Rutherford, W. (1987). *Second Language Grammar: Learning and teaching*, London: Longman.

Rutherford, W. & Sharwood Smith, M. (1986). Consciousness-raising and universal grammar. *Applied Linguistics*, 6.2: 274-281.

Rutherford, W. & Sharwood Smith, M. (eds.) (1988). *Grammar and Second Language Teaching: A Book of Readings*. Rowley, MA: Newbury House.

Ryan, S. M. & Makorova, V. (2004). Japanese learners of English: myths and reality. In Makarova, V. & Rogers, T. (eds.) *English Language Teaching: the case of Japan*, 49-64. München : LINCOM Europa.

Salaberry, M. (1997). The role of input and output practice in second language acquisition. *The Canadian Modern Language Review/La Revue canadienne des langues vivantes*, 53.2: 422-451.

Schmidt, R. (1990). The role of consciousness in second language learning. *Applied Linguistics*, 11.2: 129-158.

Schmidt, R. (1992). Psychological mechanisms underlying second language fluency. *Studies in Second Language Acquisition*, 14.4: 357-385.

Schmidt, R. (1994a). Deconstructing consciousness in search of useful definitions for applied linguistics. *AILA Review*, 11: 11-26.

Schmidt, R. (1994b). Implicit learning and the cognitive unconscious: of artificial grammars and SLA. In Ellis, N. (ed.) *Implicit and explicit learning of languages*, 165-209. London: Academic Press.

Schumann, J. (1979). The acquisition of English negation by speakers of Spanish: a review of the literature. In Andersen, R. (ed.) *The acquisition and use of Spanish and English as first and second languages*, 1-32. Washington, DC: TESOL

Sharwood Smith, M. (1981). Consciousness-raising and the second language learner. *Applied Linguistics*, 2.2: 159-169.

Sharwood Smith, M. (1991). Speaking to many minds: on the relevance of different types of language information for the L2 learner. *Second Language Research*, 7.2: 118-132.

Sharwood Smith, M. (1993). Input enhancement in instructed SLA. *Studies in Second Language Acquisition*, 15: 165-179.

Sheen, R. (1994). A critical analysis of the advocacy of the task-based syllabus. *TESOL Quarterly*, 28.1: 127-151.

Sheen, R. (2002) A response to Lightbown's (2000) 'anniversary article: classroom SLA research and second language teaching'. *Applied Linguistics*, 23.4: 519-528.

Sheen, R. (2005). Focus on forms as a means of improving accurate oral production. In House, A & Pierrard, M. (eds.). *Investigations in Instructed Second Language Acquisition*, 271-310. Berlin: Mouton de Gruyter

Sheen, Y. (2004). Corrective feedback and learner uptake in communicative classrooms across instructional settings. *Language Teaching Research*, 8.3: 263-300.

Silver, R. E. (1999). Learning conditions and learning outcomes for second language acquisition: input, output, and negotiation. Unpublished doctoral dissertation. Pennsylvania: The University of Pennsylvania, Philadelphia.

Silver, R. E. (2000). Input, output and negotiation: conditions for second language development. In Swierzbin, B., Morris, F., Anderson, M., Klee, C. & Tarone, E. (Eds.) *Social and cognitive factors in second language acquisition: selected proceedings of the 1999 Second Language Research Forum*, 345-371. Somerville, MA: Cascadilla Press.

Skehan, P. (1998). *A Cognitive Approach to Language Learning*. Oxford: OUP.

Smith, R. C. & Imura, M. (2004). Lessons from the past: traditions and reforms. In Makarova. V. & Rogers, T. (eds.) *English Language Teaching: the case of Japan*, 29-48. München : LINCOM Europa.

Spada, N. & Lightbown, P.M. (1993). Instruction and the development of questions in L2 classrooms. *Studies in Second Language Acquisition*, 15.2: 205-224.

Spada, N. & Lightbown, P.M. (1999). Instruction, first language influence, and developmental readiness in second language acquisition. *The Modern Language Journal*, 83.1: 1-22.

Storch, N. (1998). A classroom-based study: insights from a collaborative reconstruction tasks. *ELT Journal*, 52.4: 291-300.

Storch, N. (1999). Are two heads better than one? Pair work and grammatical accuracy. *System*, 27.3: 363-374.

Storch, N. (2002). Patterns of interaction in ESL pair work. *Language Learning*, 52.1: 119-158.

Swain, M. (1984). A review of immersion education in Canada: research and evaluation studies. In *A collection of U.S. Educators*. California State Department of Education.

Swain, M. (1985). Communicative competence: some roles of comprehensible input and comprehensible output in its development. In Gass, S. & Madden, C. G (eds.) *Input in Second Language Acquisition*, 235-253. Rowley, MA: Newbury House.

Swain, M. (1995). Three functions of output in second language learning. In Cook, G & Seidlhofer, B. (eds.) *Principle and Practice in Applied Linguistics. studies in honour of H. G Widdowson*, 125-144. Oxford: OUP.

Swain, M. (1998). Focus on form through conscious reflection. In Doughty, C & Williams, J. (eds.) *Focus on Form in Classroom Second Language Acquisition*, 64-82. Cambridge: CUP.

Swain, M. (2000). French immersion research in Canada: recent contributions to SLA and Applied Linguistics. *Annual Review of Applied Linguistics*, 20: 199-212.

Swain, M. & Lapkin, S. (1995). Problems in output and the cognitive processes they generate: a step towards second language learning. *Applied Linguistics*, 16.3: 371-391.

Swain, M. & Lapkin, S. (1998). Interaction and second language learning: two adolescent French immersion students working together. *Modern Language Journal*, 82:3: 320-337.

Swain, M. & Lapkin, S. (2000). Task-based second language learning: the uses of the first language. *Language Teaching Research*, 4.3: 251-274.

Swain, M. & Lapkin, S. (2001). Focus on form through collaborative dialogue: exploring task effects. In Bygate, M., Skehan, P., Swain, M. (eds.) *Researching Pedagogic Tasks: Second Language Learning, Teaching and Testing*. Harlow, Essex: Longman.

Swan, M. (2005). Legislation by hypothesis: the caseof task-based instruction. *Applied Linguistics*, 26.3: 376-401.

Taguchi, N. (2002). Implementing oral commuImplementing Oral Communication Classes in Upper Secondary Schools: a Case Study. *The Language Teacher*, 26 .12: 3-8.

Takahashi, E. (1998). Language development in social interaction: a longitudinal study of a Japanese FLES program from a vygotskyan approach. *Foreign Language Annals*, 31.3: 392-406.

Tarone, E. (2000). Still wrestling with 'context' in interlanguage theory. *Annual Review of Applied Linguistics*, 20: 182-198.

Tin, T. B. (2003). Does talking with peers help learning? the role of expertise and talk in convergent group discussion tasks. *Journal of English for Academic Purposes*, 2.1: 53-66.

Tomlin, R. S. & Villa, V. (1994). Attention in cognitive science and second language acquisition. *Studies in Second Language Acquisition*, 16.2: 183-203.

Tsang, W. (2004). Feedback and uptake in teacher-student interaction: an analysis of 18 English lessons in Hong Kong secondary classrooms. *Regional Language Centre Journal*, .35: 187-209.

van Lier, L. (1988). *The Classroom and the Language Learner*. London: Longman.

van Lier, L. (1996). *Interaction in the Language Curriculum: Awareness, Autonomy, and Authenticity*. London: Longman

van Lier, L. (2000). From input to affordance: Social-interactional learning from an ecological perspective. In Lantolf, J. *Sociocultural Theory and Second Language Learning*, 245-260. Oxford: OUP.

VanPatten, B. (1993). Grammar teaching for the acquisition-rich classroom. *Foreign Language Annals*, 26.4: 435-450.

VanPatten, B. (1994). Evaluating the role of consciousness in second language acquisition: Terms, linguistic features & research methodology. *AILA Review*, 11.11: 27-36.

VanPatten, B. (1996). *Input Processing and Grammar Instruction: Theory and Research*. Norwood, NJ: Ablex.

VanPatten, B. (2002). Processing instruction: An update. *Language Learning*, 52.4: 755-803.

VanPatten, B. & Cadierno, T. (1993). Explicit instruction and input processing. *Studies in Second Language Acquisition*, 15.2: 225-243.

VanPatten, B. & Oikkenon, S. (1996). Explanation versus structured input in processing instruction. *Studies in Second Language Acquisition*, 18.4: 495-510.

Varonis, E. & Gass, S. (1985). Non-native/non-native conversations: a model for negotiation meaning. *Applied Linguistics*, 6.1: 71-90.

Vygotsky, L. S. (1978). *Mind and Society*. Cambridge: MA: Harvard University Press.

Vygotsky, L. S. (1981). The genesis of higher mental functions. In Wertsch (ed.) *The Concept of Activity in Soviet Psychology*. Armonk, NY: M. E. Sharpe.

Vygotsky, L. S. (1987). *The Collected Works of L. S. Vygotsky. Volume 1. Thinking and Speaking*. New York, NY: Plenum Press.

Wada, M. (1989). Speech given at the Kansi Mid-Year Block Conference in Osaka, December 19889 (cited by McConnell 2000:52).

Wajnryb, R. (1990). *Grammar Dictation*. Oxford: OUP.

Watcyn-Jones, P. & Hoard-Willams, D. (1995). *Grammar 1: games and activities*. London: Longman.

Weir, C. (1990). *Communicative language testing*. Hemel Hempstead: Prentice Hall.

Wertsch, J. V. (1985). *Vygotsky and the Social Formation of Mind.* Cambridge, MA: Harvard University Press.

White, L. (1987). Against comprehensible input: the input hypothesis and the development of second language competence. *Applied Linguistics*, 8.2: 95-110.

White, L. (1991). Second language competence versus second language performance: UG or processing strategies? In Eubank, L (eds.) *Point counterpoint: Universal Grammar in the second language*, 133-161. Amsterdam: John Benjamins

White, L. (1998). Getting the learners' attention : a typographical input enhancement study. In Doughty, C & Williams, J. (eds.) *Focus on Form in Classroom Second Language Acquisition*, 85-113. Cambridge: CUP.

White, L., Spada, N., Lightbown, P. M. & Ranta, L. (1991). Input enhancement and L2 question formation. *Applied Linguistics*, 12.4: 416-432.

Wode, H. (1976). Developmental sequences in naturalistic L2 acquisition. *Working Papers on Bilingualism*, 11: 1-13.

Zobl, H. (1983). Markedness and the projection problem. *Language Learning*, 33.3: 293-313.

Zobl, H. (1985). Grammars in search of input and intake. In Gass, S. and Madden, C. G. (eds.). *Input in Second Language Acquisition*, 329-344. Rowley, Mass.: Newbury House.

Game

'GUESS WHO?'. Middlesex: Hasbro International Inc.

CD-ROMs

Art Explosion. (1996). California: Nova Development Corporation.